

DATA 512: Part 2 - Extension Plan

Introduction

In part 1 of the project, we laid the foundation for understanding the potential impact of smoke from wildfires on the Madison community, located in Dane County, WI. We utilized the past wildfire data to estimate how much smoke has reached the city in recent years and compared with AQI values for Madison city to validate our smoke estimates. Using this information, we built a predictive model to predict smoke patterns that Madison might experience over the next 25 years. In Part 2, the extension plan, we aim to expand upon this analysis, diving deep into the impact of smoke on public health and healthcare systems in Madison.

The goal of this extension plan is to help inform the city council, city manager/mayor, and city residents about the potential future impacts of smoke on their community's health and healthcare systems. We aim to particularly study the impact of smoke on respiratory illness and mortality rates in Madison over the next 25 years. In the following sections of this document, we will delve deeply into the problem statement driving this analysis, with particular emphasis on the critical role healthcare plays in our study and explain the relevance of the selected data sources. Additionally, we will identify any unknown factors or dependencies that could impact the results of the analysis. Finally, we will lay out a clear, step-by-step timeline for completing this extension, giving an overview of key tasks and deadlines.

Motivation/Problem Statement

This analysis is driven by the increasing health risks tied to wildfire smoke exposure, particularly for chronic respiratory conditions like COPD, asthma, and tuberculosis. As wildfires become more frequent and intense due to climate change and environmental factors, communities such as Madison are experiencing health impacts that elevate healthcare demands, costs, and risks to residents' quality of life. Our main objective is to confirm the relationship between wildfire smoke and respiratory illnesses in Madison by tracking hospitalizations and emergency department visits due to asthma, COPD, and tuberculosis from 2000 to 2023.

From a scientific perspective, understanding the health consequences of wildfire-induced smoke provides crucial insights into the direct impact of climate-related events on respiratory health. Practically, these findings can guide public health interventions, informing policies that protect vulnerable populations, especially during peak wildfire seasons. I hypothesize that the increasing levels of smoke exposure correlate with a rise in respiratory illnesses, hospitalizations, and potentially mortality rates. If confirmed, I expect these trends to continue upward if smoke levels persist or increase in coming years. Based on the insights uncovered and the severity of any effects observed, we can then advise the Madison city council, city manager/mayor, and city residents on potential solutions to mitigate these health impacts.

Impact Focus

The focus of this analysis is the healthcare sector of Madison, i.e. understanding the impact of wildfire smoke on respiratory health outcomes in Madison, with a particular emphasis on chronic conditions like COPD, asthma, and tuberculosis. By examining the rates of hospitalizations, emergency department visits, and mortality over the past 2-3 decades, I aim to identify whether increased smoke exposure correlates with adverse health effects in these areas. The specific research questions include:

1. Does wildfire smoke significantly affect respiratory health (COPD, asthma, and tuberculosis) outcomes in Madison?
2. Is there a relationship between wildfire smoke exposure and increased death rates from respiratory illnesses, particularly those linked to COPD, asthma, and tuberculosis?
3. What is the demand for hospitalization and emergency visits for asthma, COPD, TB in the upcoming years and how can healthcare providers prepare for this demand?

Data

We will utilize the data and model that we built in part 1 - common analysis of this project.

1. Data pertaining to Wildland Fires, provided by the USGS, are [publicly available](#) and not subject to copyright restrictions. However they can be cited as below.
“Welty, J.L., and Jeffries, M.I., 2021, Combined wildland fire datasets for the United States and certain territories, 1800s-Present: U.S. Geological Survey data release, <https://doi.org/10.5066/P9ZXGFY3>”.
2. AQI data accessed through the EPA API lies in the [public domain](#) and is not subject to domestic copyright protection under [17 U.S.C. § 105](#).
3. A polynomial regression model built on smoke estimates utilizing the above data sources. This predictive model predicts the smoke estimates for the next 25 years.

Additional Data

To enhance my analysis on respiratory illnesses, I plan to use several data sources that collectively cover mortality and morbidity data related to asthma, chronic obstructive pulmonary disease (COPD), and tuberculosis (TB). These datasets provide vital statistics and insights on hospitalization rate, emergency department visits, and mortality rates supporting a comprehensive view of respiratory illness patterns and their public health impact.

1. Asthma Data

- **Source:** Wisconsin Department of Health Services - [Asthma Data - Wisconsin DHS](#)
- **Description:** The asthma dataset includes information on county wise asthma-related hospitalizations and emergency department visits across Wisconsin from 2000 - 2023. This information is generated based on the data originally sourced from Wisconsin Hospital Association Information Center and National Center for Health Statistics.

- **Fields of interest:**

Fips	County	Topic	Year	Count	Crude Rate
5 digit fips code of the county	Count Name	Hospitalization or Emergency Visit	Cases recorded year	Number of cases	The number of cases divided by the total population of interest.

2. COPD Data

- **Source:** Wisconsin Department of Health Services - [COPD Data - Wisconsin DHS](#)
- **Description:** The COPD dataset includes information on county wise COPD-related hospitalizations and emergency department visits across Wisconsin from 2000 - 2023.
- **Fields of interest:**

Fips	County	Topic	Year	Count	Crude Rate
5 digit fips code of the county	Count Name	Hospitalization or Emergency Visit	Cases recorded year	Number of cases	The number of cases divided by the total population of interest.

3. Tuberculosis Data

- **Source:** Wisconsin Department of Health Services - [TB Data - Wisconsin DHS](#)
- **Description:** This dataset tracks the tuberculosis cases in Wisconsin from 2014 to 2023.
- **Fields of interest:**

County	Year	Count
Count Name	Year	Number of TB cases recorded

License: The above mentioned 3 datasets are publicly available for strictly non-commercial use. But users must attribute the Wisconsin Department of Health Services as the data source .

4. Mortality Data

- **Source:** Institute for Health Metrics and Evaluation (IHME) - [GBD Results Tool](#)
- **Description:** This dataset provides extensive data on mortality rates attributed to a variety of respiratory illnesses, including asthma, COPD, and TB, segmented by factors such as age and sex for the Wisconsin region for period 1980 - 2021.
- **Fields of interest:**

location_name	sex_name	age_name	cause_name	metric_name	year	val
State Name	Male / Female / Both	All Ages	Respiratory illness name	Number or Percentage	Year	Rate of mortality or count of mortality based on metric

License: [Free for non-commercial research purposes](https://vizhub.healthdata.org/gbd-results/) but must be cited as - *“Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2021 (GBD 2021) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2022. <https://vizhub.healthdata.org/gbd-results/>.”*

Using these datasets, I aim to conduct an analysis of respiratory illnesses by focusing on mortality, morbidity, and temporal variances. Combining the healthcare data along with existing smoke data and model based on year allows for exploring correlations between the wildfire smoke and respiratory health and predicting the future illness rates, possible mortality rates due to respiratory illness related to smoke and hospitalization demands.

Unknowns And Dependencies

1. Correlations observed between respiratory illnesses (such as asthma, COPD, and TB) and smoke / air quality may not imply causation. Confounding variables that could significantly impact health outcomes, such as access to healthcare, socioeconomic status, lifestyle factors, and pre-existing health conditions, are not accounted for in this analysis. Without controlling for these variables, it's challenging to isolate the effects of smoke on respiratory illnesses, which means that our conclusions should be interpreted cautiously.
2. While the analysis relies on reputable data sources, potential inaccuracies or biases in these datasets could impact the reliability of our results. Factors such as errors in data entry, limitations in sampling methods, or underreporting in certain demographic groups may lead to skewed insights. We proceed with the assumption of data integrity, yet we remain mindful of these potential limitations and interpret our findings with an awareness of their possible impact on result reliability.
3. Despite having access to smoke data spanning the last 60 years, a significant limitation is the relatively shorter temporal range of healthcare data, particularly hospitalization records, which are available only from 2000 to 2023 for asthma and COPD, and from 2014 - 2023 for TB. The lack of overlapping historical data limits our ability to examine the health impacts of smoke exposure with the same depth as our smoke data. Ideally, aligning both datasets temporally would have enabled a more comprehensive analysis.
4. County-specific mortality rate data is unavailable, compelling us to use aggregate data for the entire state of Wisconsin. By relying on statewide data, we risk obscuring local trends and nuances in mortality that might otherwise reveal significant insights. This limitation could lead to correlation errors and impact the precision of our findings regarding the health effects of smoke specifically in Madison, Dane county.
5. The development of policy recommendations is contingent on the accuracy of our model's estimates. Our polynomial regression model's predictive power relies heavily on the assumption that future trends will align with historical data, which may not account for unforeseen changes, for example another pandemic, other kind of natural catastrophe thus affecting the model's accuracy.

Timeline To Completion

For the successful completion of this extension analysis, here is a comprehensive timeline with key milestones that I have planned.

Milestones	Significant Tasks	Timeline
Data Collection and Processing	<ul style="list-style-type: none">Collect and clean mortality, and hospitalization data.Preprocess the data (Handle missing data and outliers) for consistency and accuracy.Merge these datasets with the existing dataset.	Nov 6 - Nov 10
Model Integration, Training, Validation, and Forecast	<ul style="list-style-type: none">Train the model with existing and newly added features, tune hyperparameters and validate the model using appropriate metrics.Explore alternative time series models such ARIMA if necessary.	Nov 11 - Nov 16
Visualization and Analysis	<ul style="list-style-type: none">Analyze the results to identify significant trends, patterns, and relationships.Create visualizations to communicate findings effectively.	Nov 17 - Nov 22
Presentation	<ul style="list-style-type: none">Prepare the “PechaKucha” style presentation.Incorporate visualizations, key results, and forecasts/impacts into the presentation.	Nov 23 - Nov 26
Documentation	<ul style="list-style-type: none">Ensure all code, models, project process and visualizations are properly documented.Format and organize the project on the GitHub repository as per the instructions.	Nov 27 - Dec 03

References

- Wisconsin Department of Health Services (<https://www.dhs.wisconsin.gov/>)
- The Institute for Health Metrics and Evaluation (<https://www.healthdata.org/>)
- Public Health Madison and Dane County (<https://publichealthmdc.com/health-services/respiratory-illness/dashboard>)
- U.S Centers for Disease Control and Prevention (<https://www.cdc.gov/>)
- Wisconsin Hospital Association Information Center (<https://www.whainfocenter.com/>)