Wildfire Smoke Analysis & the Impact on Health in Arlington, Texas

Raagul Nagendran

Analysis - https://github.com/raaguln/wildfire-analysis/

Introduction

There has been an alarming rate of increase in the frequency and severity of wildfires in the U.S. This has made us all realize that there is a pressing need to understand their broader impacts on community health, especially in urban areas where people may face health risks due to dense populations and limited mitigation options. Wildfire smoke, which releases harmful pollutants like carbon monoxide (CO), particulate matter, and smoke compounds, degrades air quality, leading to respiratory and cardiovascular health issues in humans.

The motivation behind this analysis is that by finding out more about these patterns between wildfires and their resultant health conditions, we could make informed decisions for emergencies, health aid allocation, and public health announcements during wildfire events. By finding patterns between mortality data and carbon monoxide exposure trends with wildfire data, the analysis seeks to provide a more nuanced view of how air quality degradation due to wildfires translates into measurable health outcomes. This could be valuable for local authorities, who could use the findings to implement more targeted interventions during wildfire episodes, and healthcare providers, who might anticipate and manage unusual increases in wildfire-related health issues.

We will focus on the healthcare section in Arlington, Texas, and its neighboring area - Tarrant County. We will specifically focus on the possible issues that could arise due to the smoke and particulate pollutants released in wildfires. This will involve deaths caused due to respiratory diseases in and around Arlington, including the whole county of Tarrant (where the city is located). Another analysis we will be doing will involve studying the relationship between people being exposed to carbon monoxide (calls made to the Texas Poison Center Network). Understanding these impacts is critical, as it highlights how wildfire smoke might cause severe health issues, which puts additional strain on healthcare systems and emergency services.

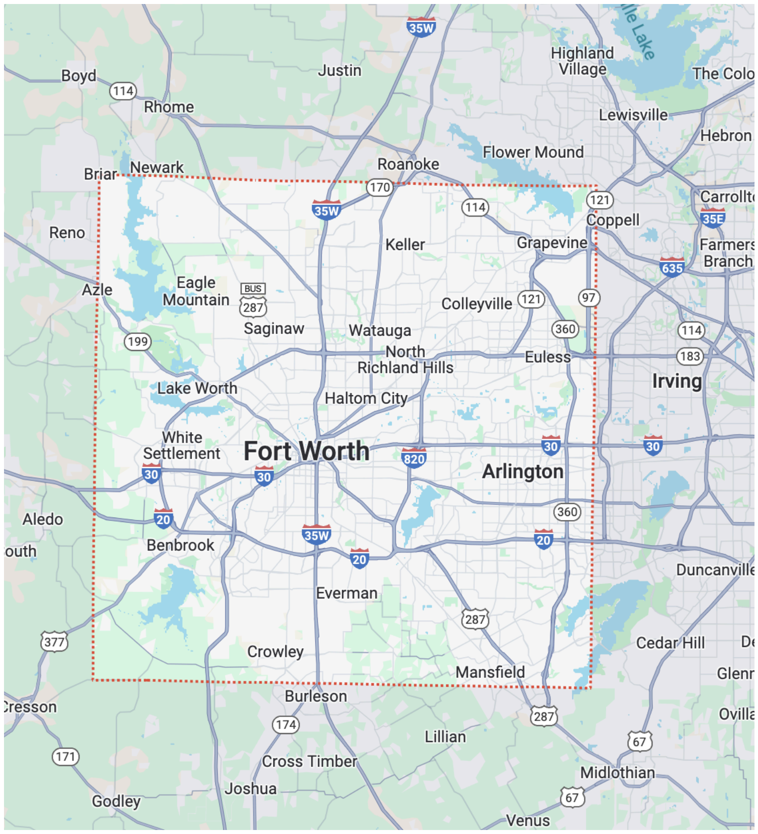

Figure 1: Arlington, Texas in the county of Tarrant (highlighted in red dotted lines)

Background/Related work

Studies have shown that even short-term exposure to wildfire smoke can result in significant increases in hospital admissions for respiratory and cardiovascular issues. For instance, a study by Liu et al. (2017) found a 7.2% increase in respiratory admissions during smoke wave days with high wildfire-specific PM2.5 levels compared to non-smoke days, indicating that short-term spikes in air pollution due to wildfire smoke can worsen existing health conditions. Moreover, prolonged exposure has been linked to higher mortality rates, particularly during intense wildfire seasons. Research also indicates that cumulative short-term exposures can lead to a reduction in lung function and may contribute to premature death, especially among vulnerable populations such as the elderly. The U.S. EPA (2020) has noted that continuous exposure over multiple days can have serious health implications, including increased risk of death from respiratory and cardiovascular diseases.

Carbon monoxide (CO) poisoning presents another critical health risk, especially during and after wildfire events. CO is an odorless, colorless gas that can lead to severe health consequences, including neurological damage and death, when inhaled in high concentrations. Over the past several decades, research has established a direct link between wildfire smoke and adverse air quality, primarily due to fine particulate matter (PM2.5) and other toxic pollutants released during combustion. These pollutants include carbon monoxide (CO), nitrogen oxides (NOx), and volatile organic compounds (VOCs), which exacerbate respiratory and cardiovascular health conditions. While Arlington, Texas, is not traditionally considered a wildfire-prone area, the influence of distant wildfires on its air quality underscores the far-reaching impacts of wildfire smoke, carried by atmospheric currents over hundreds or even thousands of miles.

Regarding the model used for forecasting, ARIMAX (AutoRegressive Integrated Moving Average with eXogenous variables) was chosen for forecasting in this analysis because it is well-suited for time series data where past values and external factors influence future outcomes. Unlike linear regression, which assumes a direct and static relationship between variables, ARIMAX accounts for temporal dependencies by considering patterns such as trends, seasonality, and autocorrelation in the data. This is critical for accurately predicting how wildfire smoke impacts health metrics over time, as these relationships often evolve and depend on historical data.

Additionally, ARIMAX incorporates exogenous variables—factors outside the time series itself—to improve predictions. In our case, variables like smoke estimates, average wildfire distance, and GIS-derived burn areas act as predictors that influence air quality and health outcomes. While other models like machine learning-based regressions or exponential smoothing methods could be used, they either require significantly larger datasets or fail to capture the nuanced time-dependent effects as effectively as ARIMAX. Therefore, ARIMAX provides a balance of interpretability and predictive power, making it an ideal choice for this scenario where both historical trends and external variables drive the forecast.

Methodology

The methodology for this analysis was structured into five distinct stages, each stage tackling a different challenge, but building towards our deeper understanding of wildfires, smoke exposure, and their related health outcomes.

This analysis was done in a way that ensured that all legal limits on data usage were strictly obeyed, as well as all license agreements for data use, as this analysis was done for a non-profit cause and in an open-source manner. Most importantly, it was created with the strongest ethical concerns in mind, ensuring that no personally identifying information was utilized (the data sources used were anonymized at the source level).

Throughout the analysis, Pearson's correlation analysis was used to understand the link between the different factors that might affect our smoke analysis. This method was chosen because it is effective at determining the strength and direction of associations between two continuous variables.

Data cleaning and preprocessing

The first step involved bringing all relevant data sources into a usable format. This included acquiring datasets from wildfire incident reports, air quality monitors, and health records. Different data cleaning and processing were applied to address issues such as missing entries and inconsistencies, followed by merging datasets to create datasets of uniform format that would facilitate easy analysis. This step ensured that the subsequent analyses were built on reliable and cohesive information. More information on each step has been documented in detail in the code notebooks and can be found in detail in the linked repository.

Calculating smoke estimates

The second step focused on quantifying wildfire intensity and calculating the smoke estimates. To find the impact of wildfire smoke on Arlington, Texas, we developed a formula that uses key factors that might affect smoke production and dispersion. First, we filtered the wildfire dataset to include only fires within 650 miles of Arlington, as fires beyond this distance are unlikely to contribute significantly to local air quality. This filtering step ensures that our analysis focuses on relevant wildfire events. From the available data, we identified three key variables critical for calculating the smoke estimate: wildfire intensity, burn area, and distance from the city.

The wildfire intensity was derived from the *Assigned_Fire_Type* column, which classifies fires based on their nature (e.g., Wildfire, Prescribed Fire). These nominal values were encoded to reflect their relative contribution to smoke production. The burn area was represented by the *GIS_Acres* column, which captures the total area affected by the fire. Since *GIS_Acres* provides a direct measure of fire size, it was selected since it was more reliable than length or width, which primarily describe shape rather than extent. The average distance was calculated by measuring the distance between Arlington and each wildfire's location. The smoke impact was assumed to decrease with increasing distance, making this variable inversely proportional in the formula.

Using these variables, the smoke estimate formula was defined as:

Smoke estimate = (Intensity of fire * Area of wildfire) / Average distance from the city

This formula brings out the fire-specific characteristics and spatial relationships to model smoke impact accurately. The analysis was restricted to data from 1960 to 2020, aligning with the most recent 60 years of wildfire data available. This range of years provides sufficient historical context while focusing on data relevant to modern environmental conditions.

Forecasting smoke estimates

In the third stage, future trends in smoke exposure were predicted using an ARIMAX (AutoRegressive Integrated Moving Average with Exogenous variables) time series model. This model leveraged historical smoke estimates along with external predictors such as weather patterns and wildfire activity to forecast future levels of smoke. This model was chosen for its ability to incorporate additional variables and provide robust predictions. More details on the reason for choosing this model are mentioned in depth in the Backgrounds section.

The fourth step involved quantifying the health impacts of smoke exposure by examining its correlation with health outcomes, such as respiratory deaths and carbon monoxide poisoning. Correlation tests, statistical analyses, and visual comparisons using plots and bar graphs were conducted to uncover relationships between smoke estimates and health metrics, revealing demographic-level disparities and other patterns. This step also ended up highlighting external influences, such as harsh winters in Arlington, which led us to the hidden additional complexity of the analysis and helped us understand the findings better.

Finally, the insights from the analysis were used to develop actionable recommendations. These included strategies for improving air quality monitoring, enhancing public awareness, and implementing targeted health interventions to mitigate the impact of wildfires on vulnerable populations. These suggestions are explained in depth in the Discussions and Limitations section.

Findings

Initial Findings

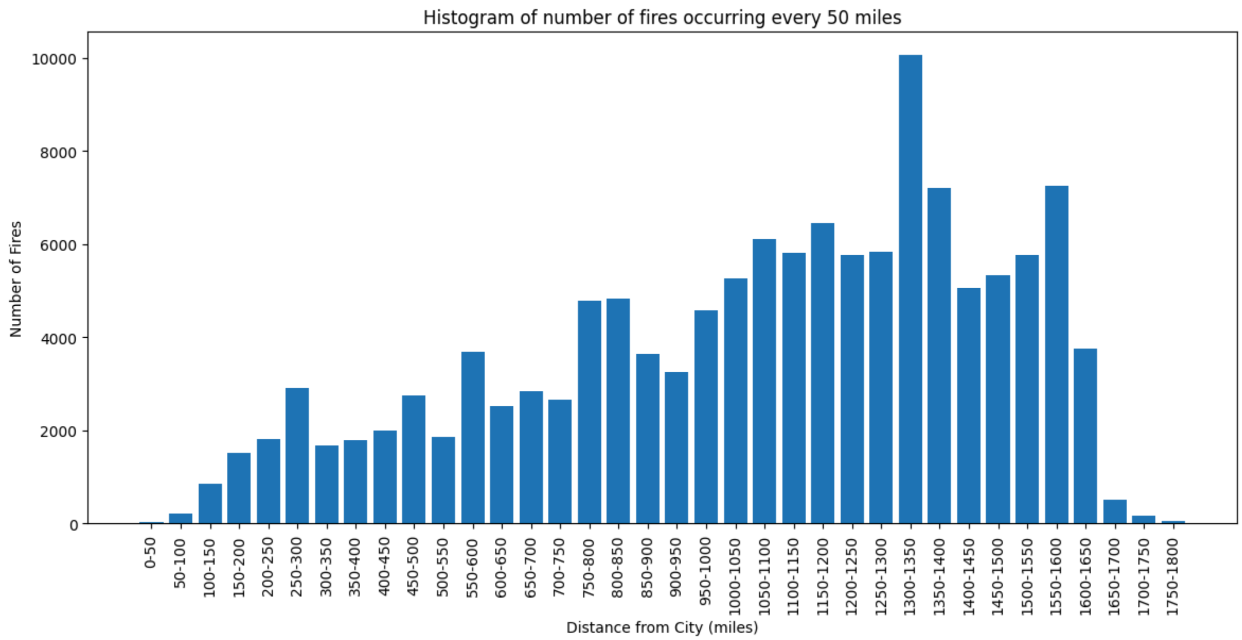

Figure 2: Histogram showing the number of fires occurring every 50 miles from the city of Arlington, Texas.

The histogram (Figure 2) displays the number of wildfires that occurred at varying distances from Arlington, Texas, grouped into bins of 50 miles each. The x-axis represents the distance from the city in miles, while the y-axis indicates the number of fires recorded in each distance range. The histogram shows that the majority of wildfires occur between 1000 and 1350 miles away, with the highest peak around 1300–1350 miles. In contrast, very few wildfires are recorded within 0–100 miles or beyond 1650 miles. This distribution highlights that most wildfires impacting Arlington's air quality originate from distant regions rather than areas close to the city. This plot shows us how even distant wildfires have an impact on the air quality of our city.

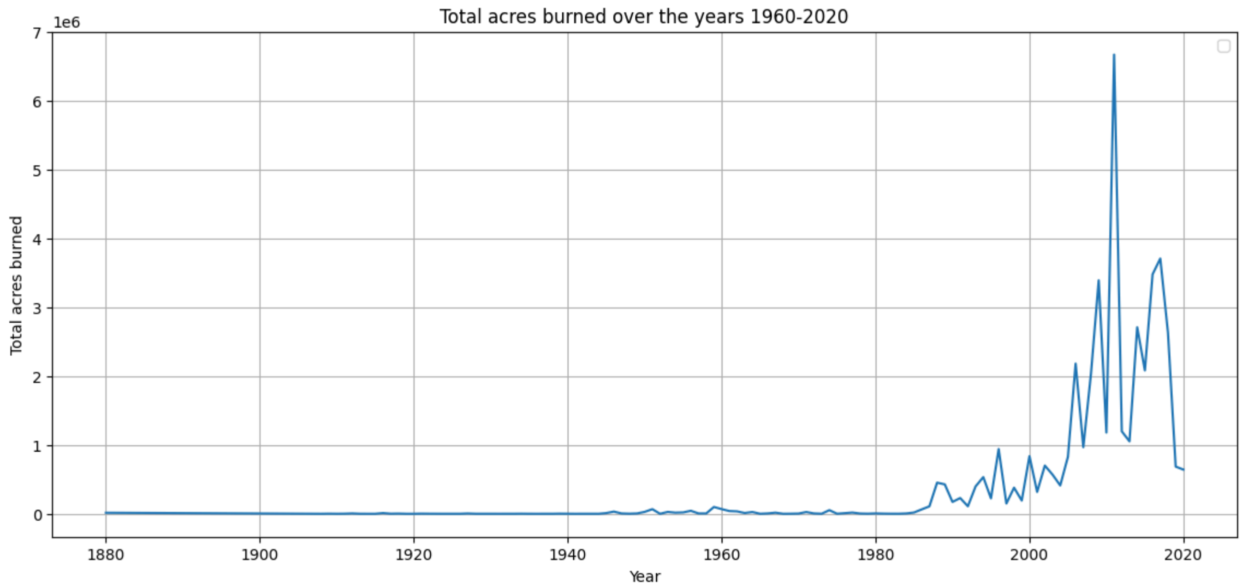

Figure 3: Line chart showing the total acres burned over the years 1960-2020

The line chart (Figure 3) shows the total acres burned by wildfires annually from 1960 to 2020. The x-axis represents the year, while the y-axis measures the total area burned in acres (normalized scale). The data indicates a significant increase in wildfire activity, particularly from the 1980s onwards. For much of the earlier period, total acres burned remained relatively low and stable. However, starting around the 1980s, there was a noticeable upward trend, with several peaks occurring in the late 1990s and 2000s, culminating in the highest recorded peak around the early 2000s, where total burned acreage exceeded 6 million. This sharp rise aligns with the growing concerns about climate change, increased drought conditions, and land-use changes, which might be contributing to more frequent and severe wildfires.

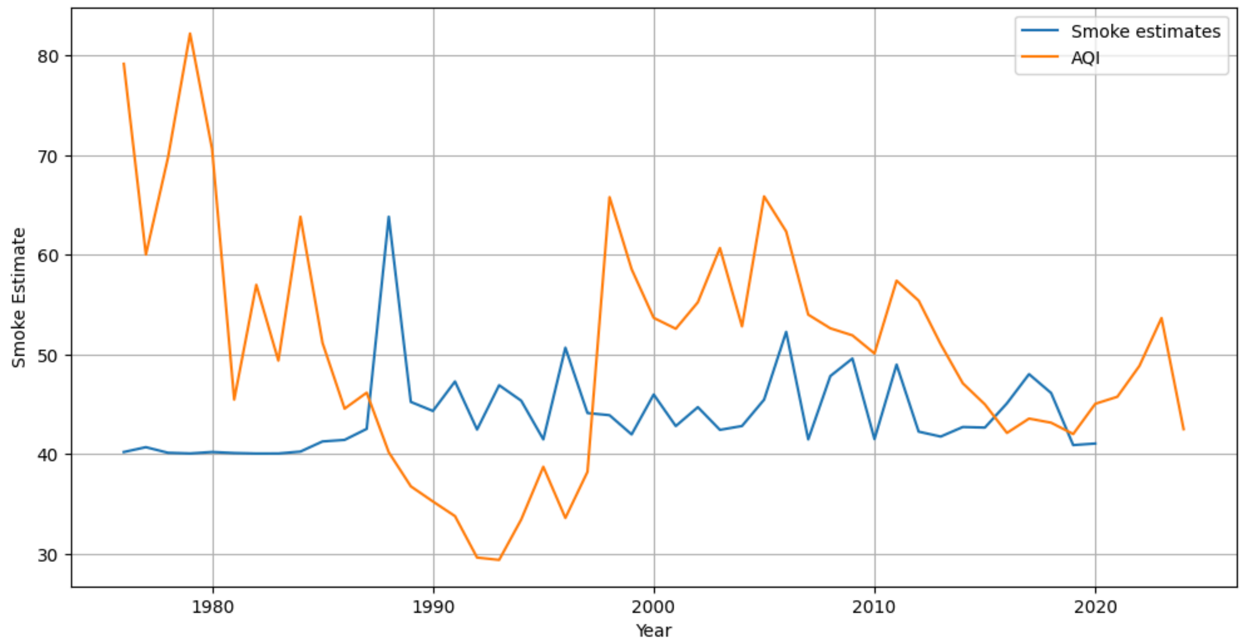

Figure 4: Smoke Estimates and AQI over the years

The line chart (Figure 4) compares two metrics, Smoke Estimate and Air Quality Index (AQI), over time, with the x-axis representing years and the y-axis representing their respective normalized values. The blue line shows the Smoke Estimate, while the orange line represents AQI. Initially, AQI values were higher in the late 1970s and early 1980s but showed a significant decline after 1990, reflecting improvements in air quality regulations or reduced pollutant emissions. The Smoke Estimate remains relatively stable, with some fluctuations, particularly after 1990, where it shows an inverse trend compared to AQI. This suggests that although smoke from wildfires contributes to air quality degradation, other factors may also influence AQI. The relationship between the two metrics becomes less correlated after 2000, with the Smoke Estimate maintaining steadier levels compared to the declining AQI. This pattern highlights that while wildfire smoke impacts air quality, other contributors to AQI, like industrial emissions or urban pollutants, may play a role in its variability. There might also be a delayed effect that might be happening, which needs to be analyzed further.

Forecasts for Smoke Estimates

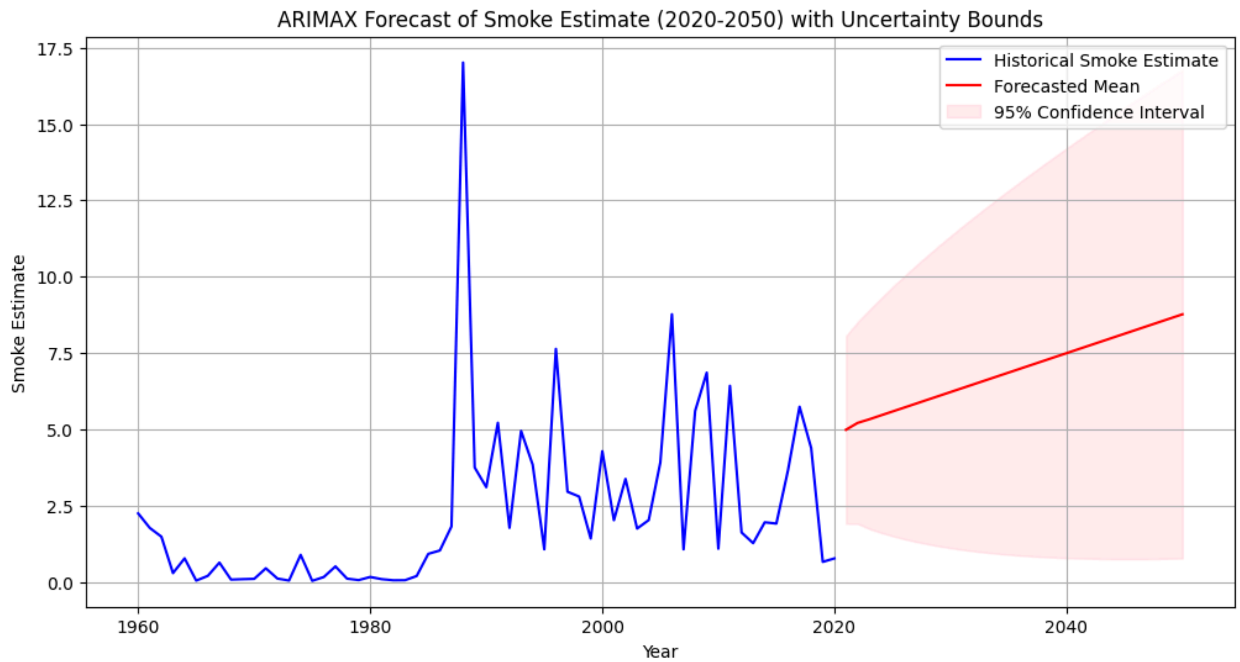

Figure 5: ARIMAX Forecast of smoke estimates for the years 2020 to 2050

The plot shows that smoke levels around Arlington, Texas, are likely to worsen if wildfires continue. The blue line represents past data, showing how smoke levels have changed over the years. While there are ups and downs, there has been a general increase since 1980. The red line shows the forecast, predicting a steady rise in smoke levels from 2020 to 2050, with the model indicating a consistent upward trend. The pink shaded area represents the confidence interval, which shows the range where actual smoke levels are expected to fall. The wider this area, the less certain the model is about its predictions, which is normal for statistical forecasts.

Findings on Health Impacts due to wildfires

The analysis (Figure 6) reveals notable patterns in the correlation between wildfire exposure and health outcomes, with demographic-level differences highlighting that certain demographic groups are more likely to have a higher level of adverse effects than others. For respiratory-related deaths, a moderate correlation of 0.24 can be seen overall, indicating a statistically significant slight link between long-term exposure to poor air quality from wildfires and increased mortality rates. Interestingly, the correlation is higher for females compared to males, suggesting that women may be more susceptible to developing respiratory complications due to wildfire smoke.

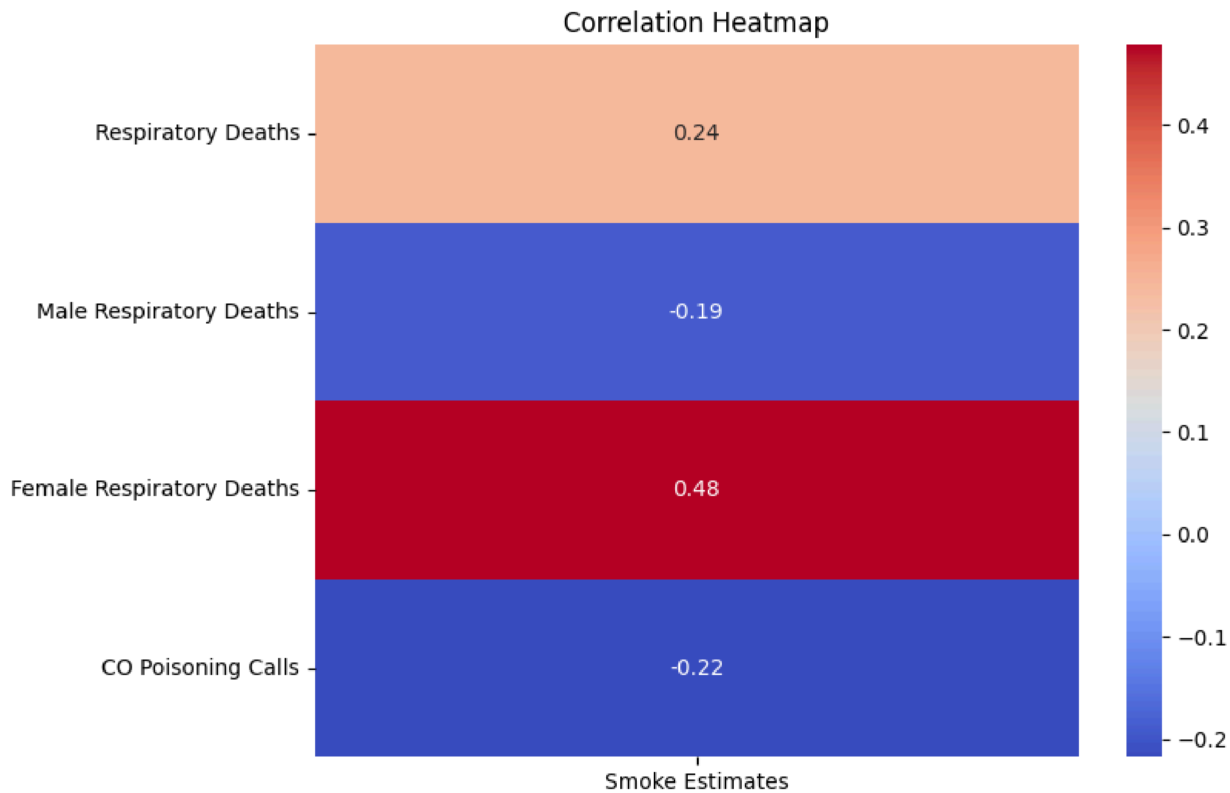

Figure 6: Correlation heatmap for the different health metrics considered

For carbon monoxide (CO) poisoning, a correlation of 0.22 was found, highlighting the short-term health risks associated with wildfires. However, plotting the data on a line chart (Figure 7) revealed an anomaly - CO poisoning calls have steadily increased since 2016, even as smoke estimates decreased during the same period. This divergence suggests that there might be factors beyond wildfire smoke that may be contributing to CO poisoning, prompting further investigation. External research points to the role of harsher winters and the increased use of unsafe indoor heating, compounded by a lack of smoke alarm mandates, as significant contributors to this rise. This is explained more in-depth in the Limitations section of the report.

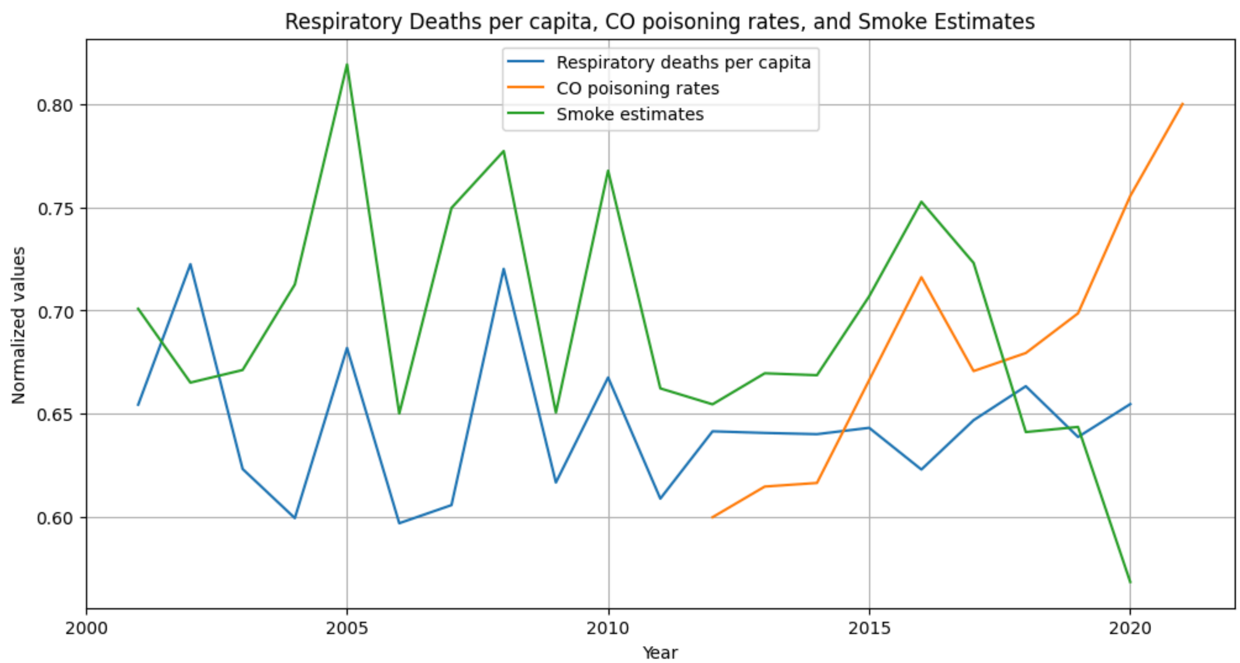

Figure 7: Respiratory deaths, CO poisoning rates, and smoke estimates, visualized over the years

Forecasts for Health Impacts due to wildfires

Looking to the future (Figure 8), projections indicate that health impacts from wildfires are likely to worsen if current trends persist. The blue line represents past data, showing how each health-related fatality has changed/increased over the years. The red line shows the forecast, predicting a steady rise in fatality for all 3 health-related issues from 2020 to 2050. The pink shaded area represents the confidence interval, which shows the range where actual fatality in the health issues is expected to fall. The wider this area, the less certain the model is about its predictions, which is normal for statistical forecasts.

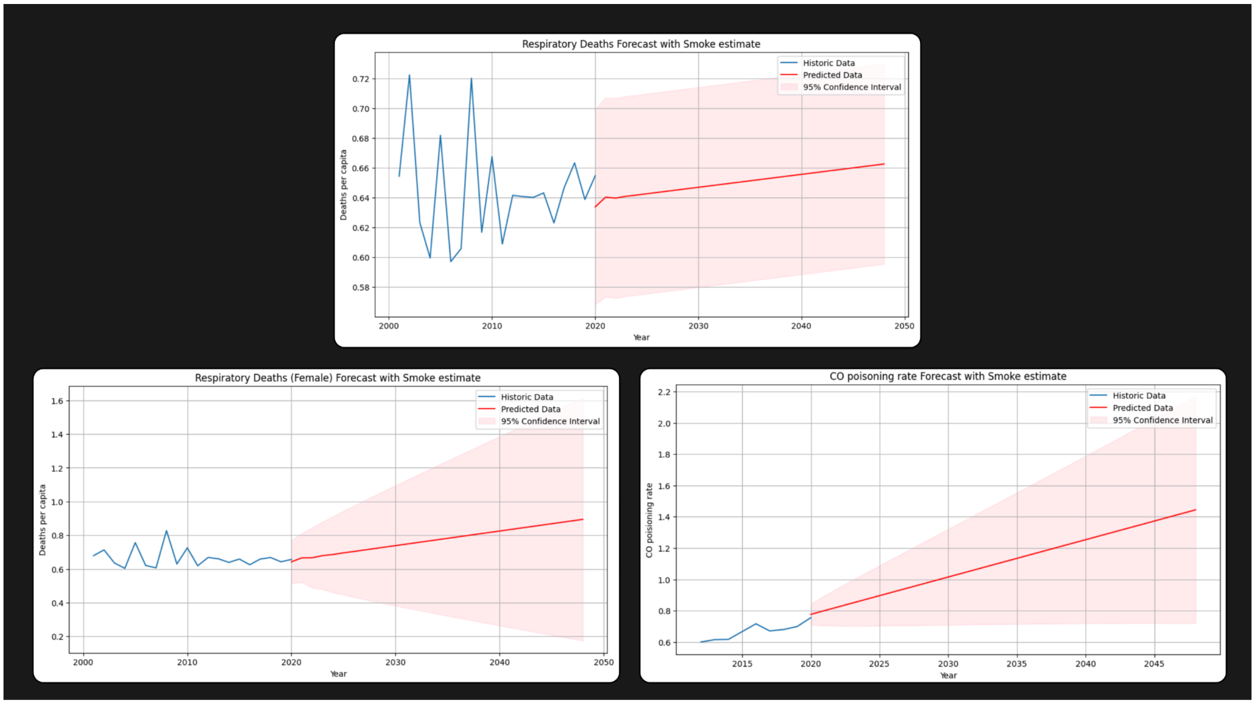

Figure 8: Forecasting health impacts of wildfire smoke

With increasing wildfire activity due to climate change and other contributing factors, the burden of respiratory diseases and CO poisoning could grow, especially among vulnerable populations. This highlights the urgency of implementing preventive measures, such as improved air quality monitoring, public education campaigns, and policies to mitigate the health effects of both wildfires and other confounding influences.

Overall, the findings emphasize the multifaceted nature of wildfire-related health impacts and the importance of considering both direct and indirect factors in future research and policy-making. Addressing these challenges requires a comprehensive approach that integrates environmental, health, and socioeconomic data to better protect communities from the growing threat of wildfires.

Discussions

The findings of this analysis are significant as they provide a comprehensive understanding of how distant wildfires, increasing wildfire activity, and smoke exposure contribute to air quality degradation and public health risks in Arlington, Texas. Most notably, the data reveals that wildfires occurring over 1,000 miles away can still impact local air quality, emphasizing the importance of adopting a regional and even national approach to wildfire management. These insights are particularly important for informing policies and community planning, as they demonstrate the interconnected nature of environmental events and local health outcomes. The alarming rise in total acres burned over recent decades, coupled with projections of worsening smoke exposure and health impacts, calls for urgent action to mitigate these threats.

Local governments should advocate for stronger collaboration with state and federal agencies to address wildfire management and air quality control. Specifically, policies that promote controlled burns, improved forest management, and investments in wildfire suppression technologies are critical. At the local level, the city council should consider expanding air quality monitoring systems and establishing early warning systems for high-smoke days. Public health campaigns can educate residents on mitigating wildfire-related health risks, such as using HEPA filters and wearing masks during smoke events. Given the increasing frequency and intensity of wildfires, stakeholders have a limited window—likely 3 to 5 years—to develop and implement a concrete, long-term plan to address these challenges effectively.

There are other similar strategies that the city can take to ensure reduced impact of wildfires. Effective wildfire prevention relies heavily on proactive forest management strategies, such as controlled burns, thinning dense vegetation, and removing dead trees. By reducing fuel loads in fire-prone areas, these measures help to minimize the intensity and spread of wildfires. Public education also plays a critical role in creating awareness about the risks of wildfires. Evacuation plans and personal protective measures are important, such as using air purifiers or wearing masks during smoke events. Educating residents on the signs of wildfire danger and how to stay informed can save lives and reduce health risks associated with smoke exposure.

This project looks at wildfires through a human-focused lens. It understands that environmental disasters hit vulnerable people the hardest. The researchers want to create solutions that help entire communities. Their approach goes beyond just technical methods like tracking air quality or managing forests. They also think about teaching people, running health programs, and getting community members involved. The goal is to create helpful plans that everyone can understand and use. The researchers work to make sure their recommendations are fair and give people tools to protect themselves. They talk to many different groups, from local leaders to everyday citizens, to make sure their plans work for everyone.

Limitations

Careful background work and research were done to ensure that there is no extrapolation of the results beyond what the data was indicating. Despite this, this analysis of wildfire impacts on health faces several limitations. A key challenge is the influence of confounding variables, such as changing weather patterns, industrial pollution, and socioeconomic factors, which independently affect air quality and health outcomes. For example, while wildfire smoke is a significant contributor to air quality degradation, other sources like pollution emissions from everyday city activities and indoor pollutants complicate the impact of health outcomes directly to wildfires. These confounding factors introduce uncertainty into the analysis, making it difficult to come to definitive causal relationships between wildfire activity and the above-mentioned deaths and poisonings.

One unique limitation specific to Arlington, Texas, is the abnormal rise in CO poisoning since 2016. The abnormal increase in CO poisoning post-2016 needed looking into, so a lot of external background research was done to understand if this was due to some external factors that go beyond the wildfires. The research led us to a definitive reason behind the increase in CO poisoning. This increase was due to harsher winters, which led to more widespread use of unsafe indoor heating methods, such as burning charcoal or using unventilated heaters. What further increased this issue to its abnormal increase and peaks is the lack of a city or state mandate requiring smoke alarms in homes, as it always is warm in Texas, even during winters. Without this critical safety measure, CO poisoning incidents started going undetected until they became severe, resulting in higher poisoning rates. This unique situation highlighted how different environmental and human factors complicate the analysis where we set out to find the health impacts of wildfire smoke. This strongly emphasizes the need for policy interventions, such as mandating smoke alarms and public safety education campaigns.

Data quality issues also present a significant limitation to the study. Wildfire data may suffer from inaccuracies, such as the misclassification of prescribed fires as wildfires or the underreporting of smaller fire events, which can distort exposure estimates. Similarly, health data, including respiratory disease-related deaths and CO poisoning records, may contain inconsistencies due to underreporting, incomplete datasets, or reliance on secondary sources. These limitations introduce biases and gaps in the data, reducing the reliability of the findings and the ability to establish precise linkages between wildfire activity, air quality, and health outcomes.

Finally, the scope of this study is limited by its focus on a subset of contributing factors and specific health outcomes. While it provides valuable insights into the relationship between wildfires, air quality, and health, it does not account for indirect impacts, such as economic losses, psychological stress, or long-term chronic illnesses. Addressing these limitations would require more comprehensive data collection methods, including integrating environmental, health, and socioeconomic datasets, and employing advanced statistical techniques to account for confounding factors. Future studies should also explore broader impacts to better inform holistic policy responses and community resilience strategies.

Conclusions

We started this analysis to examine how wildfires, even those occurring far from Arlington, Texas, influence local air quality and public health outcomes. The findings ended up revealing several key insights: distant wildfires, primarily those over 1,000 miles away, are significant contributors to air quality degradation in Arlington. The analysis also highlights serious and alarming trends in wildfire activity. The total acres burned has risen sharply since the 1980s, a pattern tied to climate change and land-use changes. These findings show us how interconnected regional and national wildfire events are with local environmental and health outcomes, signaling the need for coordinated action across geographic boundaries.

Through the lens of human-centered data science, this research highlights the disproportionate health impacts of wildfire smoke on vulnerable populations, like women. But this might also be a false alarm, which can only be confirmed by doing a much more comprehensive analysis of all the socio-economic and health factors related to a society's presence, which is way beyond the scope of this project. But if there is any truth to this discovery, the study not only deepens the understanding of wildfire-related health disparities among the different social groups but also identifies critical areas for targeted interventions. For instance, public health initiatives could make use of targeted strategies to address these disparities by providing education on protective measures for women.

Apart from the obvious immediate findings, this analysis emphasizes the need for diverse strategies that span different aspects of our societal life to address the wildfire challenges comprehensively. The rise in CO poisoning in Arlington since 2016, driven by harsher winters and unsafe indoor heating practices, highlights the complex interplay of environmental and policy factors. This emphasizes the urgent need for systemic changes, such as mandating smoke alarms and developing community resilience strategies. Overall, the study advances the understanding of wildfire impacts and demonstrates the value of human-centered data science in translating complex data into actionable insights that prioritize community well-being and resilience.

References

1. Liu, J. C., Wilson, A., Mickley, L. J., Dominici, F., Ebisu, K., Wang, Y., Sulprizio, M. P., Peng, R. D., Yue, X., Son, J.-Y., & Anderson, G. B. (2016). Critical Review of Health Impacts of Wildfire Smoke Exposure. *PMC*, *28*(1), 77–85. https://doi.org/10.1097/EDE.0000000000000556
2. U.S. Environmental Protection Agency. (n.d.). Health Effects Attributed to Wildfire Smoke. Retrieved December 5, 2024, from https://www.epa.gov/wildfire-smoke-course/health-effects-attributed-wildfire-smoke
3. Welty, J.L., & Jeffries, M.I. (2021). *Combined wildland fire datasets for the United States and certain territories, 1800s-Present*. U.S. Geological Survey. Retrieved from https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81
4. U.S. Environmental Protection Agency. (n.d.). *Frequent questions about AirData*. Retrieved from https://www.epa.gov/outdoor-air-quality-data/frequent-questions-about-airdata
5. Centers for Disease Control and Prevention. (n.d.). *Underlying Cause of Death, 1999-2020*. WONDER Online Database. Retrieved from https://wonder.cdc.gov/mcd.html
6. Texas Department of State Health Services. (n.d.). *Carbon Monoxide Related Poison Calls*. Health Data. Retrieved from https://healthdata.dshs.texas.gov/dashboard/environmental-health/carbon-monoxide-related-poison-calls
7. Tarrant County Public Health. (n.d.). *Chronic Disease Data and Information*. Retrieved from https://www.tarrantcountytx.gov/en/public-health/public-health-administration/health-data-and-information/chronic-disease.html?linklocation=Public%20Health%20Data%20and%20Information&linkname=Chronic%20Disease%3C/a
8. North Carolina State University. (2021, December). *Explainer: How Wildfires Start and Spread*. Retrieved from https://cnr.ncsu.edu/news/2021/12/explainer-how-wildfires-start-and-spread/
9. Texas Department of State Health Services. (n.d.). *Carbon monoxide-related poison center calls*. Retrieved from

https://healthdata.dshs.texas.gov/dashboard/environmental-health/carbon-monoxide-related-poison-calls

10. Trevizo, P. (2021, August 19). *Inside the 911 response to carbon monoxide poisoning in Texas*. The Texas Tribune. Retrieved from https://www.texastribune.org/2021/08/19/texas-winter-storm-carbon-monoxide-poisoning/
11. Dyer, O. (2021, February 16). *Hospitals see a rise in hypothermia cases, and carbon monoxide poisonings as Texans struggle to stay warm*. Dallas Morning News. Retrieved from https://www.dallasnews.com/news/weather/2021/02/16/hospitals-see-rise-in-hypothermia-cases-carbon-monoxide-poisonings-as-texans-struggle-to-stay-warm/
12. Trevizo, P., Larson, R., Churchill, L., Hixenbaugh, M., & Khimm, S. (2021, April 29). *Texas power outages led to a carbon monoxide poisoning catastrophe*. The Texas Tribune. Retrieved from https://www.texastribune.org/2021/04/29/texas-carbon-monoxide-poisoning/
13. Gibbons, J. (2016, February 7). *Family of six 'dropped in their spots' from carbon monoxide poisoning*. MLive. Retrieved from https://www.mlive.com/news/flint/2016/02/family_of_six_dropped_in_their.html
14. Texas Poison Center Network. (n.d.). *Texas Poison Center Network*. Retrieved from https://www.poisoncontrol.org/home/

Data Sources

1. **USGS**: Provides comprehensive data on wildfire locations and characteristics, including fire perimeter, size, and duration. This information is crucial for understanding wildfire impacts and trends.
2. **EPA**: Offers air quality data, including the Air Quality Index (AQI), which tracks pollutants like ozone, particulate matter, and carbon monoxide. This dataset helps assess air quality and its effects on public health.
3. **CDC Wonder**: A resource for understanding respiratory disease-related deaths. It provides access to a variety of health-related data, allowing users to analyze trends and causes of respiratory diseases.
4. **Texas Poison Center Network**: Contains data related to carbon monoxide (CO) poisoning incidents in Texas. This information is essential for public health initiatives and understanding the prevalence of poisoning cases.