Introduction

Odessa, Texas, located at the center of the Permian Basin, is central to an oil-rich region shaping West Texas's economic and industrial landscape. The Permian Basin, the most prolific oil field in the U.S., produces 45% of the nation's oil and ranks as the second-largest shale gas-producing region. While these industries drive economic growth, they also exact a heavy toll on the environment and public health. The region emits massive quantities of methane, a potent greenhouse gas, alongside harmful chemicals like benzene and volatile organic compounds (VOCs), contributing to climate change and poor air quality. Odessa experiences elevated levels of ground-level ozone, worsening respiratory conditions like asthma and COPD. Hydraulic fracturing (fracking) further compounds these issues, risking water contamination and adding to the Basin's significant carbon emissions.

In recent years, the environmental toll has been exacerbated by increasing wildfire smoke, driven by climate change and more frequent wildfires. Odessa's already stressed environment amplifies the health impacts of smoke, which can worsen asthma, COPD, and other respiratory conditions. This report examines wildfires from 1964–2024 and forecasts smoke trends for 2025–2050 to understand their impacts on Odessa's public health, particularly on respiratory conditions and birth defect cases.

The health risks of smoke exposure extend beyond physical symptoms to broader social and economic challenges, including increased healthcare costs, missed workdays, and reduced quality of life. This analysis aims to provide actionable data to inform policymakers, healthcare providers, and residents about the urgency of air quality management and targeted interventions. By highlighting evidence-based insights, the report seeks to guide strategies for addressing the compounded health risks from industrial pollution and wildfire smoke in Odessa.

While the report focuses on smoke exposure, it acknowledges that genetic predisposition, socioeconomic factors, and additional environmental stressors also influence health outcomes. Nevertheless, the clear link between smoke exposure and adverse health effects underscores the need for urgent action. This report serves as a roadmap for fostering a healthier, more sustainable future for Odessa and its residents.

Background/Related Work

Wildfire Historical Data Exploration

To estimate and forecast the wildfire smoke near Odessa, Texas, we first need to understand the trends of wildfires in recent years. Historical fire data spanning the past 60 years (1964–2024) were analyzed; however, it is notable that the available dataset does not include wildfire records beyond 2020. The background analysis scope covers wildfires occurring within a 650-mile radius of Odessa, Texas, with the annual fire season defined as May 1st through October 31st.

Existing research (cited in the Citation section) has outlined the importance of spatial and temporal analyses in understanding wildfire behavior and its environmental impacts. Prior studies have highlighted the geographic clustering of wildfires and seasonal variations, providing a foundation for estimating their proximity to urban centers and the resulting effects on air quality. These insights informed our approach to analyzing the locations of wildfires and the cumulative burned acres over time. Additionally, studies on smoke dispersion and air quality impacts emphasize the significance of integrating wildfire data with metrics like the Air Quality Index (AQI) to evaluate public health implications.

The first visualization (Fig. 1) displays the spatial distribution of wildfires relative to Odessa, TX.

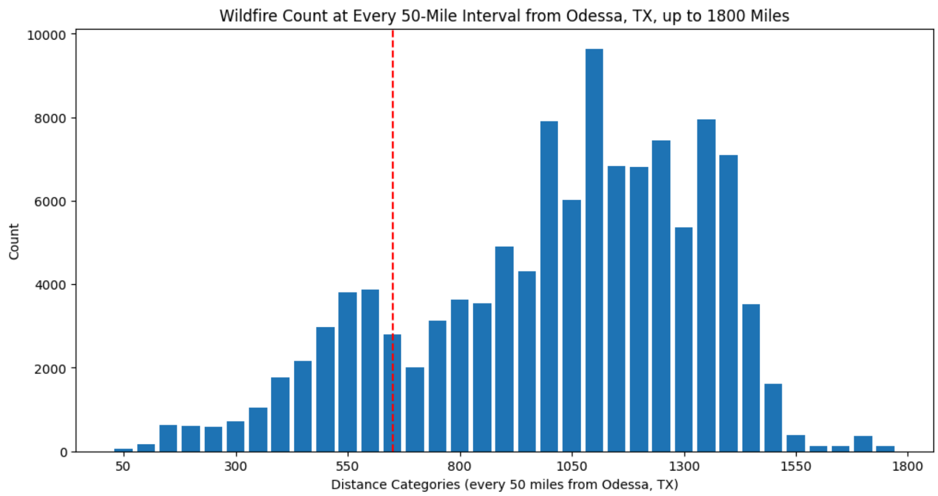


Fig. 1

Specifically, it quantifies the frequency of wildfires occurring at varying distances from Odessa within 1800 miles. The x-axis represents distance from Odessa, TX, segmented into 50-mile intervals, while the y-axis indicates the number of wildfires observed within each interval. This histogram offers insights into the geographic spread of wildfires in relation to a specific location, Odessa, TX, and helps identify if certain distances have higher wildfire occurrences than others. To construct this visualization, we began by calculating the average distance between each wildfire event and Odessa, TX. These distances were then organized into 50-mile buckets, enabling us to tally the wildfire frequency within each bucket and subsequently plot this count as a bar on the histogram. This approach allows viewers to quickly interpret wildfire proximity to Odessa and observe trends in fire occurrences across the distance spectrum.

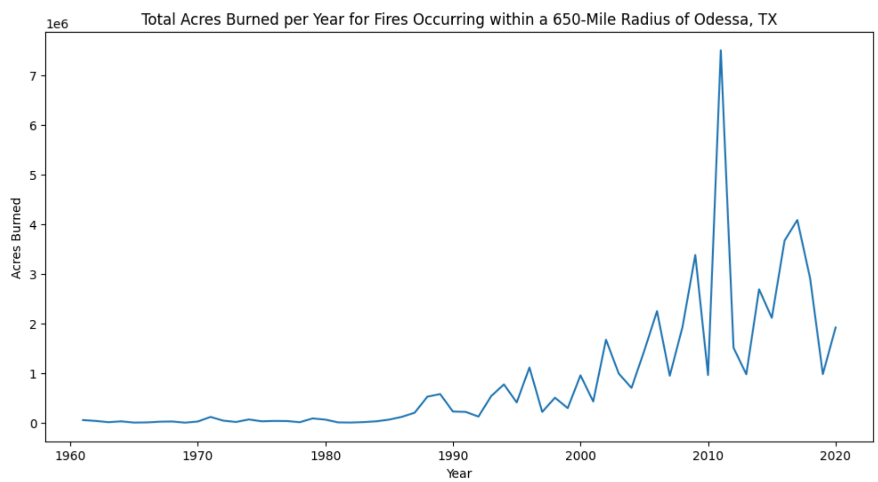


Fig 2.

The second visualization (Fig. 2) is a time series graph that illustrates the annual total acre burned by wildfires within a 650-mile radius of Odessa, TX, spanning the years 1964 to 2020 (no wildwire data after 2020). On the x-axis, each year is represented from 1964 to 2020, marking the time dimension of the data, while the y-axis quantifies the cumulative acres burned each year, offering insight into the severity and variability of fire impact over time.

To construct this graph, the wildfire data were filtered to include only those fires within 650 miles of Odessa, TX. For each year, we aggregated the total acres burned by summing the size of each individual wildfire. This time series analysis allows us to identify patterns, trends, and outliers in wildfire behavior across the decades. A notable peak occurred in 2013, with a recorded 7,504,385 acres burned, indicating an exceptionally high level of wildfire activity for that year within the specified distance from Odessa. This peak, alongside other fluctuations over the years, provides a temporal perspective on the environmental impact of wildfires near Odessa, highlighting periods of increased fire severity.

To connect wildfire activity to air quality impacts, we analyzed the Air Quality Index (AQI) for Odessa alongside estimated smoke density derived from wildfire size and proximity.

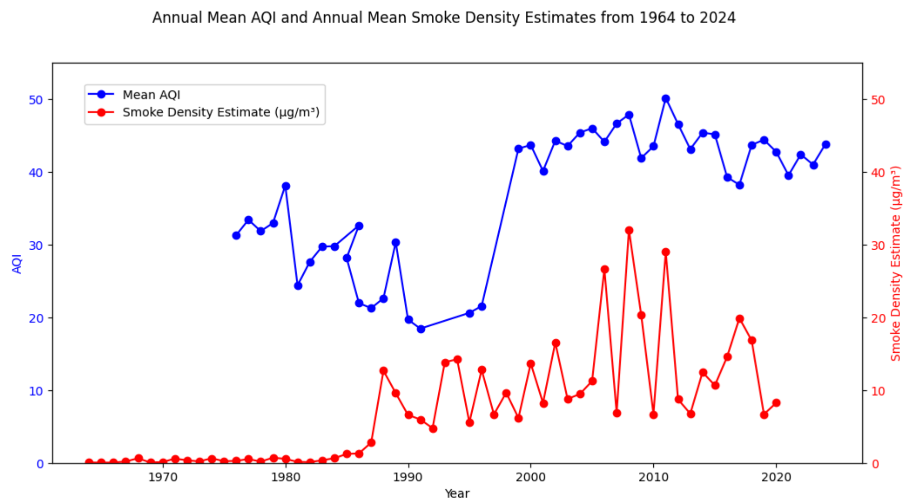


Fig. 3

The third visualization (Fig. 3) is a time series graph comparing these two metrics over various years. The x-axis represents time, while the y-axis includes scaled values for both estimated smoke density and AQI levels, facilitating direct comparison. Methods as to how the smoke is estimated will be covered in the Methodology section. While this comparison highlights correlations between wildfire activity and AQI fluctuations, the dataset contains gaps, such as missing AQI data between 1964 and 1975 and in portions of the 1990s. Despite these limitations, this visualization underscores the potential linkage between wildfire smoke and degraded air quality in Odessa.

Methods Explored for Forecasting Smoke

Two methods were explored in our attempt to forecast wildfire smoke for 2025 - 2050: linear regression and Autoregressive Integrated Moving Average (ARIMA), a statistical model used to forecast future values in time series data.

The linear regression approach assumes a direct, linear relationship between the predictor variable (year) and the response variable (smoke estimate). Using historical data, the model fits a straight line that minimizes the difference between observed and predicted smoke values. Once this trend line is established, it is extended to forecast future smoke estimates based on the historical relationship.

A primary strength of the linear regression model is its simplicity and interpretability. By using a straightforward mathematical relationship, the model is easy to understand and provides clear insights into how the year influences smoke estimates. This clarity is particularly valuable for communicating findings to a wide range of stakeholders, including policymakers and community members, who may not be familiar with more complex statistical approaches.

However, a significant limitation of the linear regression approach lies in its assumption of a strictly linear relationship between year and smoke. In reality, the relationship between time and smoke levels is far more complex and influenced by nonlinear factors, such as variations in wildfire activity, climate change, and regional environmental policies. Historical data reveal fluctuations in smoke estimates that do not align perfectly with a straight-line trend, suggesting that a linear model may oversimplify the dynamics of smoke generation and dispersion. As a result, the model fails to capture important patterns or shifts in smoke behavior over time, potentially leading to inaccurate forecasts.

Further research lends us to the model that we eventually used for forecasting smoke: The ARIMA (Autoregressive Integrated Moving Average) model, which is a well-established statistical approach used for analyzing and forecasting time series data. It has been extensively applied in various domains, including environmental science, economics, and meteorology, to predict future trends based on historical observations. ARIMA is particularly effective for datasets with temporal dependencies, making it an appropriate choice for forecasting annual wildfire smoke levels, which are influenced by trends and year-to-year variability.

Prior research has demonstrated ARIMA's strength in capturing complex temporal relationships by integrating three components:

- Autoregression (AR): Models the relationship between current and past values.
- Differencing (I): Removes non-stationary trends to stabilize the dataset.
- Moving Average (MA): Accounts for patterns in residual errors.

By leveraging historical smoke estimates and fitting an ARIMA model, we can explore how past patterns might inform future trends. This approach builds on existing work in time series forecasting while addressing the specific challenges of modeling long-term smoke impacts in a changing climate.

Additional Data Sources

To aid in understanding how smoke has historically impacted and are currently impacting Odessa, TX through the healthcare lens, we incorporate additional relevant health related data to achieve our objective. All datasets curated in the Data Sources section at the end. All data sources are linked in the Data Source section below.

Centers of Disease Control and Prevention

Compressed Mortality

The *Compressed Mortality File (CMF)* is a comprehensive national database from the Centers for Disease Control and Prevention, providing county-level data on mortality and population from 1968 to 2016. This dataset is valuable for our analysis as it allows us to examine historical trends in mortality specifically linked to respiratory diseases, which are potentially influenced by smoke exposure in Odessa, TX. For this project, we will focus on data from *"Ector County, Texas"*, isolating cases where the Cause of Death is categorized under *"Diseases of the respiratory system."* The dataset includes critical variables such as *County, Year, Deaths, Population, Crude Rate Per 100,000, Age-Adjusted Rate Per 100,000*, and *Percent of Total Deaths*, providing a robust foundation for assessing the public health impact of smoke-related respiratory issues in the region.

United States and Puerto Rico Cancer Statistics, 1999-2021 Incidence/Mortality

The *United States Cancer Statistics (USCS)* provides the official federal statistics on cancer incidence and mortality across the 50 states and the District of Columbia, sourced from registries that meet high-quality data standards. Produced by the Centers for Disease Control and Prevention (CDC) and the National Cancer Institute (NCI), the USCS database is a valuable resource for understanding cancer trends. For this project, our focus is on Texas, specifically targeting respiratory-related cancers, including those categorized under *"Respiratory System," "Nose, Nasal Cavity and Middle Ear," "Larynx," "Lung and Bronchus," "Pleura,"* and "Trachea, Mediastinum and Other Respiratory Organs." Since county-level cancer data is unavailable, we will use state-level data for our analysis.

The terms of use for both the *Compressed Mortality File* and the *United States and Puerto Rico Cancer Statistics* are as follows:

1. Use these data for statistical reporting and analysis only.
2. Do not present or publish statistics representing nine or fewer births or deaths, including rates based on counts of nine or fewer births or deaths, in figures, graphs, maps, tables, etc. Statistics representing one through nine (1-9) births or deaths are suppressed on CDC WONDER.
3. Make no attempt to learn the identity of any person or establishment included in these data.
4. Make no disclosure or other use of the identity of any person or establishment discovered inadvertently and advise the Confidentiality Officer in the National Center for Health Statistics of any such discovery.

PLACES: County Data

This dataset contains model-based, county-level estimates provided by the Centers for Disease Control and Prevention, Division of Population Health, Epidemiology and Surveillance Branch. Our analysis will focus specifically on estimates related to asthma and chronic obstructive pulmonary disease (COPD), including both the crude prevalence and age-adjusted prevalence of these conditions.

The *PLACES data* is intended for public use, primarily for statistical reporting and analysis, and the CDC encourages its application to support public health initiatives. Users of this data must adhere to standard data usage practices, ensuring that their work maintains confidentiality and does not compromise individual privacy.

Texas Birth Defects Registry (TBDR) Annual Report, 1999 - 2020

The *Texas Birth Defects Registry (TBDR) Annual Report* provides records of birth defects across various body systems in Texas, with data including year, cases, prevalence rate, and confidence interval. Although we could only access statewide records, this data will help us analyze trends in birth defect counts over time. Additionally, we will examine the distribution of different types of birth defects, as there is no clear indication of any particular defect being more prominent. This broad approach will allow us to assess potential impacts on pregnancies that may have been at risk due to smoke exposure.

The Texas Health Data Use Policy allows users to reproduce and utilize data published by the Texas Department of State Health Services (DSHS), provided that DSHS is credited as the source, the access date and publication date are noted, and copyright notices are retained without modifications to the data. DSHS does not control the end use of downloaded data, nor does it provide any warranty, guarantee, or quality assurance for analyses conducted with the data. Users must include a disclaimer in any work product stating that DSHS makes no representation, warranty, or guarantee regarding the data or analyses derived from it. This policy does not create any enforceable rights or benefits and does not limit DSHS's legal responsibilities or protections under applicable law.

Methodology

Methods to Estimate Smoke for 1964 - 2024

In addition to understanding the historical trends of wildfires and the AQI, we explored ways to estimate annual wildfire smoke. To estimate the smoke from wildfires near Odessa, TX, the following formula is developed

$$S = k * (Fire\ size_{acres}) * 0.0015625 * \frac{1}{(Distance_{miles})^2} * \alpha(1 + (Y - Y_0))$$

Variable explanations:

- S: estimated smoke density in µg/m³
- k: a coefficient representing smoke emission intensity. Tentatively set to 5 µg/m³ due to the lack of specific smoke data and to follow the AQI (shown in Fig 3)
- $Fire\ size_{acres}$: size of each wildfire, positively correlating with the amount of smoke produced
- 0.0015625: conversion factor for acres to square miles (1 square mile = 640 acres)
- $\frac{1}{(Distance_{miles})^2}$: Reflects the inverse-square law of distance, where smoke density decreases with the square of the distance from the fire source
- α: a coefficient representing how intense smoke level changes since Y_0. Tentatively set to 5 due to the lack of specific smoke data
- $Y - Y_0$: Temporal adjustment factor to account for potential variations in smoke over time. Y is the year of the wildfire and Y_0 is the year 1964 (baseline year). This term attempts to account for changes in atmospheric conditions due to global warming and increased greenhouse gases concentrations

The inverse-square law is specifically explored to understand how smoke disperses from its source. This physical principle, commonly used in fields such as acoustics and light propagation, states that the intensity of an effect (in our case, smoke density) is inversely proportional to the square of the distance of the source. This relationship is crucial for modelling smoke distribution, as it reflects that smoke intensity decreases rapidly as distance from the wildfire increases. By integrating this law into the formula, we make sure that smoke estimates accurately reflect the spatial dispersion observed in real-world scenarios.

For each wildfire, the formula was applied to estimate the smoke density (*S*). These values were then aggregated by year, calculating the mean smoke density across all wildfires occurring within a 650-mile radius of Odessa, TX. This annual mean provides an estimated measure of the region's overall smoke impact for each year. This method was used to generate the smoke density presented in Fig. 3, a time series comparing annual mean AQI and smoke density. By calculating yearly averages, the visualization highlights trends in wildfire-related smoke impacts over time.

The formula was developed with a human-centered focus on long-term impact, explainability, and ethical responsibility. The inclusion of a temporal adjustment factor accounts for climate trends and emphasizes the evolving nature of wildfire smoke risks over time, underscoring the potential long-term effects on public health. Explainability was a key consideration, as the formula was designed to be interpretable by a diverse group of stakeholders, including policymakers and community leaders. From an ethical perspective, transparency about assumptions (e.g., coefficients such as k and α) and the limitations of the data ensures that the analysis avoids misleading conclusions and supports informed decision-making.

Methods to Forecast Smoke for 2025 - 2050

Building on the historical estimation of wildfire smoke, we also aim to forecast smoke levels for 2025 - 2050. As stated in the Background/Related Work section previously, the Autoregressive Integrated Moving Average (ARIMA) model is well-suited for forecasting annual wildfire smoke levels due to its ability to handle temporal dependencies in time series data. In our implementation, the ARIMA model is configured with an order (1, 2, 1), where:

- 1 (AR) models the dependence of current values on previous observations
- 2 (I) applies differencing to remove any non-stationary trends
- 1 (MA) captures relationships between residual errors

Using this configuration, the model was fit to historical smoke density estimates using the formula presented in the previous subsection. Predictions were generated for 2025 - 2050 by extending the learned temporal relationships. The resulting forecast provides a year-by-year estimate of future smoke density, offering insights grounded in historical patterns.

Similarly to the formula developed for estimating smoke, the choice of the ARIMA model for forecasting smoke levels also reflects human-centered considerations, prioritizing relevance, transparency, and ethical responsibility. ARIMA, with its well-documented performance and flexibility, underscores a commitment to employing proven methods to produce reliable predictions. These forecasts aim to inform long-term public health planning and environmental policy development. Transparency in the method's assumptions, such as the chosen order (1, 2, 1) and reliance on historical smoke estimates, ensures that stakeholders can fully understand the reliability and limitations of the results.

From an ethical standpoint, both methods of estimation and forecasting aim to provide data-driven insights without unnecessary fear-mongering. By highlighting both historical and potential future trends, the forecast promotes awareness and resilience rather than fear. Care was also taken to acknowledge the limitations of the ARIMA forecast, such as its reliance on historical data and its inability to account for sudden changes in wildfire behavior or climate policies. This ensures that the results are interpreted with appropriate caution, empowering stakeholders to use the findings constructively in planning for the future.

Methods to Forecast Health Indicators Trends for 2025 - 2050

To forecast health-related trends for 2025 - 2050 in attempts to gain better understanding on how wildfire smoke may impact the Odessa population, we used the Vector Autoregressive Moving-Average with Exogenous Regressors (VARMAX) model. The VARMAX model captures the dynamic relationships between multiple dependent variables (endogenous variables) while incorporating the influence of external factors (exogenous variables). This makes it an ideal choice of modeling how wildfire smoke, represented by smoke density estimates, impacts a range of health indicators, such as respiratory conditions, cancer rates, and birth defects. For example, changes in *Lung and Bronchus Cancer Cases*, could directly or indirectly affect other outcomes, like *Trachea, Mediastinum and Other Respiratory Organs Cancer Cases*. By including smoke density as the exogenous variable, the model ensures that our forecasts reflect both the intrinsic relationships among health indicators and the external environmental factors driving these changes, providing a more holistic and realistic understanding of future health trends.

To effectively train the VARMAX model, we grouped the health indicators into four categories based on their interrelationships:

1. **Respiratory Conditions**: Includes asthma deaths, COPD with acute respiratory infection deaths, asthma crude prevalence, and COPD crude prevalence.
2. **Cancer Age-Adjusted Rates**: Includes age-adjusted rates for lung and bronchus cancer, respiratory system cancer, and trachea and mediastinum cancer.
3. **Cancer Case Counts**: Includes the number of cases for lung and bronchus cancer, respiratory system cancer, and trachea and mediastinum cancer.
4. **Mortality and Birth Defects**: Includes deaths related to lung and bronchus cancer, respiratory system cancer, and trachea and mediastinum cancer, as well as total birth defects cases.

For each group of health indicators, we trained a separate VARMAX model using historical data (up to 2024) as the endogenous variables and annual mean wildfire smoke density as the exogenous variable.

The following steps were used for each group:

1. **Model Training**: Historical data for the health indicators and smoke density were indexed by year and used to fit the VARMAX model. The order (1, 1) configuration was chosen to capture short-term dependencies in both the autoregressive (AR) and moving-average (MA) components.
2. **Future Smoke Estimates**: Forecasted smoke density estimates for 2025–2050 were used as the exogenous variable for generating health indicator forecasts.
3. **Forecasting**: The trained VARMAX model was used to generate forecasts for each health indicator from 2025–2050, reflecting both historical trends and the influence of future smoke exposure.

Findings

Smoke Forecast Using the ARIMA Model

As previously described, we used the ARIMA model to forecast wildfire smoke for 2025 - 2050. In Fig. 4, the blue line represents the historical smoke density data (1964–2020; limited to 2020 due to the unavailability of wildfire data beyond that year), while the red line depicts the forecasted smoke density based on the historical trend.

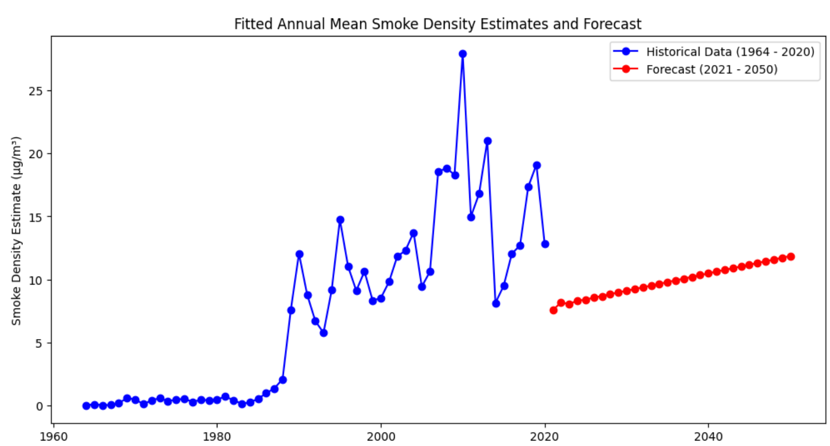


Fig. 4

The forecasted smoke density trend appears relatively linear, and this behavior can be explained by the characteristics of the ARIMA model and the nature of the historical smoke data. One key factor is the order configuration of the ARIMA model, particularly the second-order differencing ($d = 2$) applied to the data. Differencing is used to remove trends and make the data stationary, which is a requirement for ARIMA to function effectively. However, higher-order differencing, such as the second order used here, tends to smooth out variations in the data, stabilizing it. As a result, the model assumes that this stable trend will continue into the future, producing a relatively linear forecast.

Additionally, the absence of strong cyclical or seasonal patterns in the historical wildfire smoke data contributes to the linear appearance of the forecast. ARIMA models are designed to capture temporal dependencies, but without clear recurring patterns in the data, the model extrapolates the steady trend observed after differencing. Historical variability in the dataset (1964–2020) is likely treated as noise rather than meaningful signals, further reinforcing the steady continuation of the overall trend.

Health Indicators

Given the wide range of possible health indicators, we selected a set of relevant metrics that can effectively capture the potential effects of wildfire smoke. Using the additional datasets mentioned above, we analyzed indicators across three key categories:

1. **Respiratory Health**: Asthma deaths, COPD with acute lower respiratory infection deaths, asthma crude prevalence, and COPD crude prevalence.
2. **Cancer Metrics**: Age-adjusted rates, case counts, and mortality rates for lung and bronchus cancer, respiratory system cancer, and trachea, mediastinum, and other respiratory cancers.
3. **Birth Defects**: Total birth defects cases.

This spectrum of health indicators provides a comprehensive foundation for understanding the broader health impacts of smoke exposure.

Historical Trends

Fig. 5 showcases the historical trends of each of the health indicators. From top left to bottom right, the

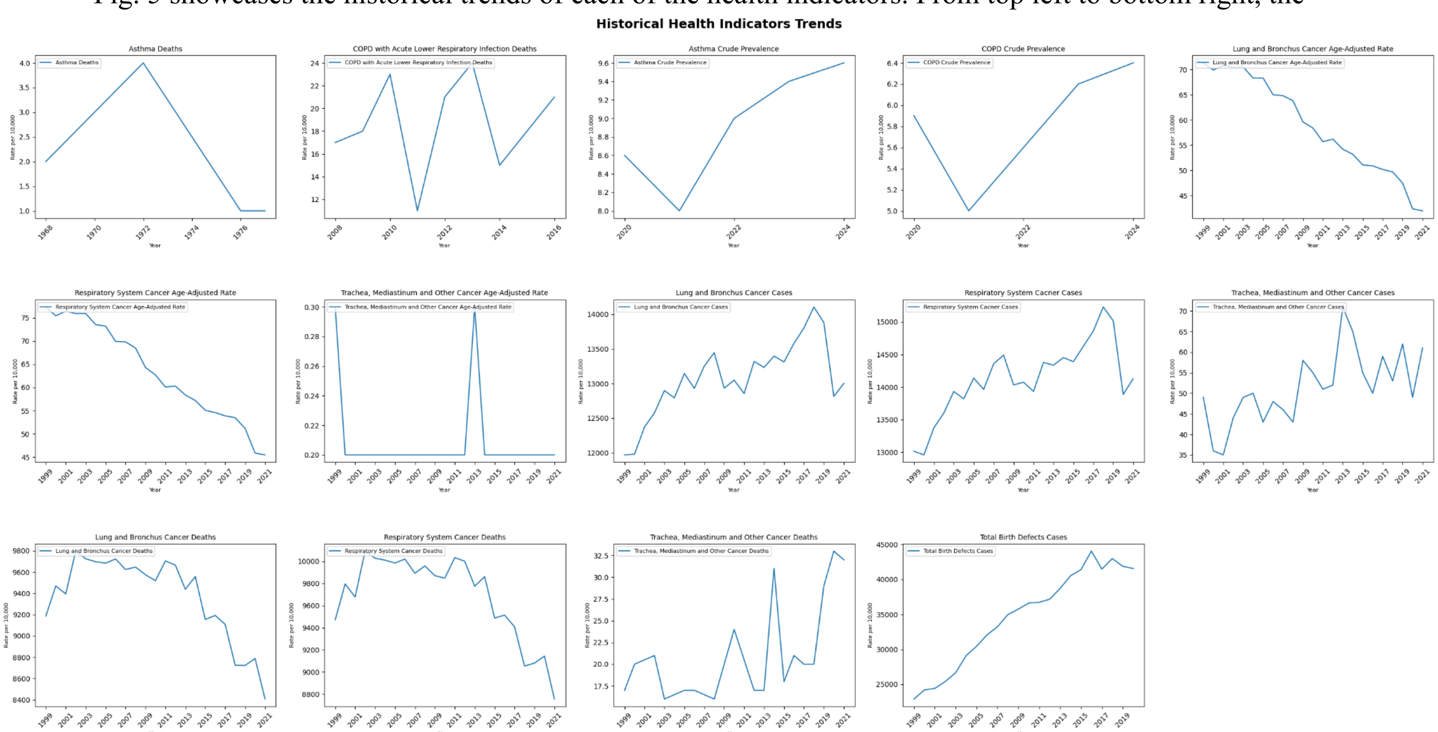


Fig. 5

subplots are *Asthma Deaths, COPD with Acute Lower Respiratory Infection Deaths, Asthma Crude Prevalence, COPD Crude Prevalence, Lung and Bronchus Cancer Age-Adjusted Rate, Respiratory*

System Cancer Age-Adjusted Rate, Trachea, Mediastinum, and Other Cancer Age-Adjusted Rate, Lung and Bronchus Cancer Cases, Respiratory System Cancer Cases, Trachea, Mediastinum, and Other Cancer Cases, Lung and Bronchus Cancer Deaths, Respiratory System Cancer Deaths, Trachea, Mediastinum, and Other Cancer Deaths, and *Total Birth Defects Cases*.
The historical trends of health indicators provide valuable insights into the evolving impacts of respiratory and cancer-related health outcomes over the years. As shown in Fig 5, the trends reveal significant variability across different indicators.

- **Respiratory Indicators**: Indicators such as asthma deaths, COPD deaths, and crude prevalence of asthma and COPD exhibit diverse trends. For instance, asthma crude prevalence shows a consistent increase in recent years, while COPD-related deaths highlight fluctuations.
- **Cancer Indicators**: Cancer-related metrics, including age-adjusted rates and case counts, exhibit notable trends. For example, lung and bronchus cancer age-adjusted rates show a gradual decline. However, the total number of cancer cases for respiratory-related cancers, such as lung and trachea, demonstrates a rising trend.
- **Birth Defects**: Total birth defect cases in Texas show a steady increase over time.

The observed patterns provide a baseline for understanding how future exposures and interventions might influence these indicators. By leveraging these insights, the study aims to forecast trends and inform public health planning for 2025–2050.

Correlation between Smoke and Health Indicators

In addition to understanding the historical health trends, we are interested in how each is correlated to the wildfire smoke estimate.

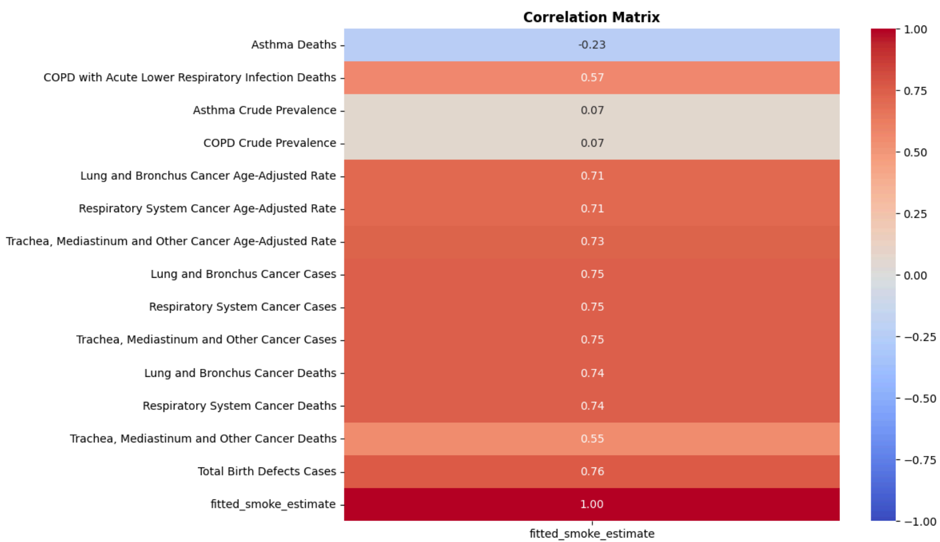


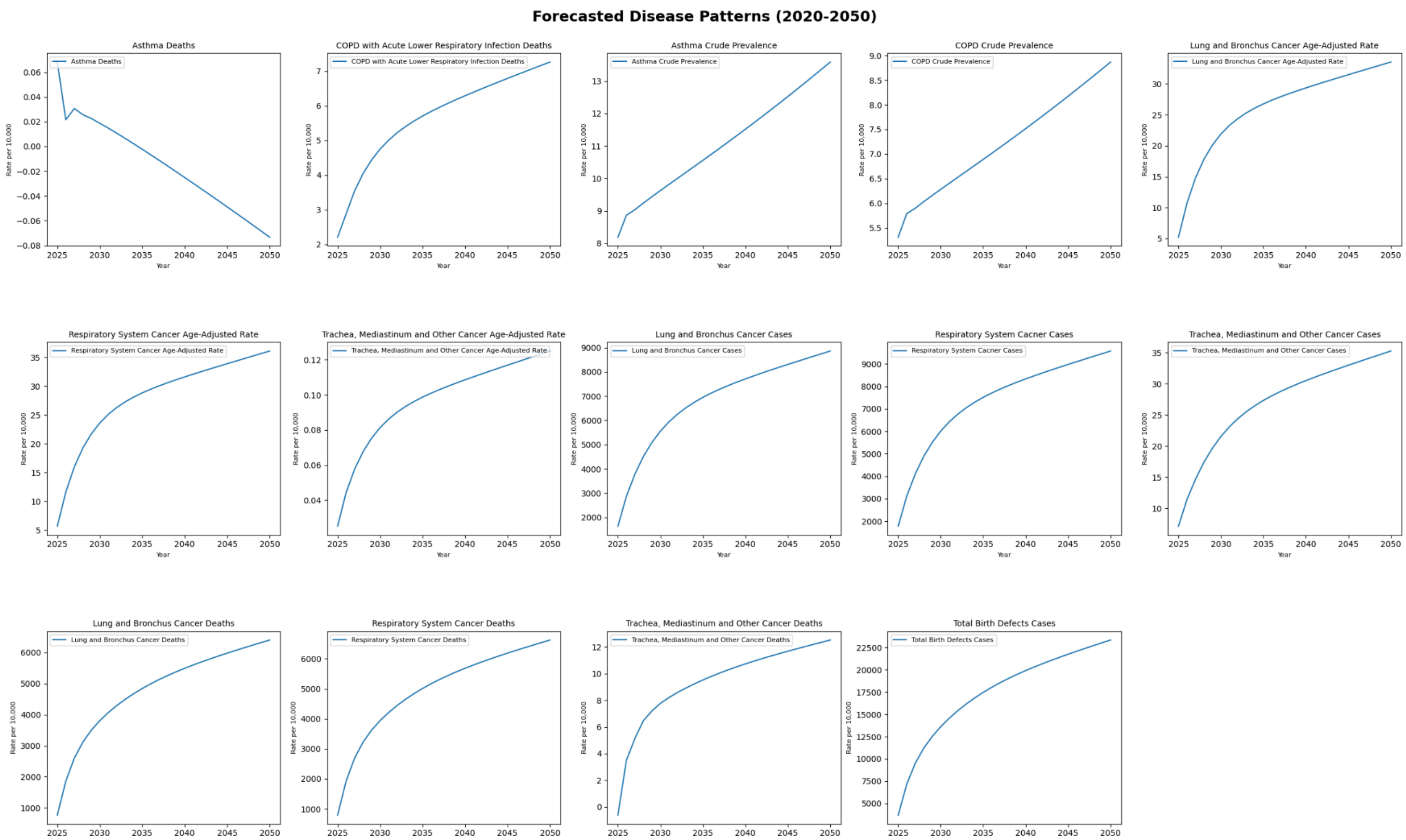
Fig. 6

The correlation matrix (Fig. 6) highlights significant relationships between fitted smoke estimates and various health indicators, particularly in cancer, respiratory conditions, and birth defects. *Total Birth Defects Cases* show the strongest correlation (0.76), suggesting a substantial association between smoke density and birth outcomes. Cancer-related indicators, such as *Lung and Bronchus Cancer Cases* (0.75) and *Respiratory System Cancer Cases* (0.75), along with their age-adjusted rates (0.71–0.73), also exhibit strong positive correlations, indicating a potential link between smoke exposure and cancer prevalence. Respiratory conditions, such as *COPD with Acute Lower Respiratory Infection Deaths*, show a moderate correlation (0.57), reflecting the impact of smoke on severe respiratory outcomes. However, *Asthma Deaths* (-0.23) and *prevalence rates for Asthma and COPD* (both 0.07) display weak to no correlations, suggesting these conditions may be more influenced by other factors. These findings emphasize the critical role of wildfire smoke in driving adverse health outcomes, particularly in vulnerable populations, while highlighting the need for further research to explore other contributing factors.

Project Trends for 2025 - 2050

To complete the health indicators analysis, we are also interested in forecasting each health indicator for the years 2025 - 2050 using the VARMAX model previously laid out. These forecasts provide insight into the potential long-term health impacts of wildfire smoke on respiratory conditions, cancer-related outcomes, and birth defects in Odessa, Texas or Texas in general. The results are shown in Fig. 7. Respiratory health indicators, such as *Asthma Deaths* and *COPD with Acute Lower Respiratory Infection Deaths,* show divergent trends. While asthma deaths are projected to decrease gradually, COPD-related deaths show a slow but increase. Similarly, prevalence indicators for *Asthma Crude Prevalence* and

Fig. 7



COPD Crude Prevalence both show linear upward trends, highlighting a growing burden of respiratory conditions in the region.

For cancer-related health outcomes, age-adjusted rates for *Lung and Bronchus Cancer*, *Respiratory System Cancer,* and *Trachea, Mediastinum, and Other Cancers* exhibit steady growth, reflecting the cumulative impact of smoke and potentially other environmental factors over time. A similar trend is observed in cancer case counts and deaths, with consistent increases forecasted for *Lung and Bronchus Cancer Cases*, *Respiratory System Cancer Cases*, and related mortality indicators. *Total Birth Defects Cases* are also projected to rise steadily.

Discussion/Implications

The findings of this analysis underscore the growing threat that wildfire smoke poses to public health, particularly in an already environmentally stressed region like Odessa, Texas. The ARIMA forecast reveals a steady increase in wildfire smoke density from 2025 to 2050, which, combined with the projected rise in respiratory conditions, cancer rates, and birth defects, highlights the urgent need for action. These findings are particularly alarming given that Odessa is situated in the Permian Basin, where industrial pollution already compounds public health risks. If left unaddressed, the combination of industrial emissions and increasing wildfire smoke could severely strain healthcare systems, reduce quality of life, and exacerbate socioeconomic disparities.

Given these risks, the city council, city manager/mayor, and city residents must take immediate steps to address these challenges. First, the city should prioritize the development of a comprehensive, long-term plan to mitigate the health impacts of smoke exposure. This plan should focus on two key areas: improving healthcare accessibility and enforcing stricter environmental regulations. Healthcare investments should include expanding access to preventive care, early detection programs for respiratory and cancer conditions, public health education initiatives, and reproductive care. A robust healthcare system will enable the community to respond more effectively to the anticipated rise in health issues. While immediate changes are ideal, a concrete plan should be established within the next 5–10 years to prevent worsening outcomes as the smoke density continues to increase.

Second, stronger wildfire prevention measures must be enacted. This includes implementing stricter regulations on land-use practices, enforcing fire safety policies, and increasing funding for forest management strategies, such as controlled burns and fuel reduction. Public education campaigns can also play a critical role in reducing human-caused wildfires and informing residents about the risks of smoke exposure.

As stated in the Methodology section, this project was guided by several human-centered data science principles to ensure the findings are both actionable and equitable. Transparency was emphasized not only through the use of interpretable models like ARIMA and VARMAX but also by fully disclosing the limitations of the models and smoke data, including reliance on historical trends and the absence of external predictors such as climate or land-use changes. Inclusivity was addressed by selecting a diverse range of health indicators to reflect the varied impacts of smoke exposure on vulnerable populations, such as children and individuals with pre-existing conditions. Ethics informed the presentation of results, focusing on empowering stakeholders with actionable insights while avoiding fear-mongering and ensuring the use of non-identifiable health data. Lastly, the principle of data literacy shaped the creation of clear visualizations and straightforward explanations, making the findings accessible to non-experts and fostering a more informed and engaged community capable of driving meaningful change.

Limitations

This project has several limitations that should be considered when interpreting the results. These span the modeling approach, data assumptions, and external factors, each with potential implications for the findings.

The smoke estimation formula is an oversimplified representation of real-world conditions, relying on assumptions such as the inverse-square law for smoke dispersion and fabricated coefficients (k and α)due to the lack of actual smoke measurements. This limits the accuracy and reliability of the smoke estimates and forecasts. The absence of critical predictors, such as wildfire type, duration, and fuel composition, further reduces the formula’s ability to capture the complexities of smoke production. Additionally, averaging smoke densities for all wildfires in a given year may not accurately reflect the true smoke impact for that year.

The ARIMA model for smoke forecasting assumes that future trends will follow historical patterns and does not account for external factors such as climate change, wildfire management policies, or land-use changes, which could cause significant deviations from the forecast. The model also lacks exogenous variables like temperature, precipitation, or vegetation changes, limiting its ability to capture the multifaceted dynamics of wildfire smoke. While ARIMA provides a useful baseline, these limitations highlight the need for complementary models incorporating non-linear trends and external predictors.

Data quality issues also impact the results, particularly in the health indicator datasets, which contain many missing values. Missing health data was handled by filling gaps with zeros, potentially underrepresenting or misrepresenting actual trends and obscuring meaningful patterns. Additionally, the smoke data was constructed using a fabricated formula rather than direct measurements, further limiting its validity. The reliance on state-level data for cancer and birth defects, due to the lack of county-level data, may not fully represent conditions specific to Odessa, Texas.

Finally, while the health indicator forecasts use smoke estimates as predictors, the VARMAX models do not account for other influential factors such as genetic predisposition, socioeconomic conditions, and unrelated environmental stressors. For instance, hereditary factors, healthcare advancements, and randomness in health outcomes contribute to the observed trends, which cannot be fully modeled. Focusing solely on smoke exposure as the primary driver may oversimplify the complex interplay of factors influencing health indicators.

In summary, these limitations emphasize the need for future work that incorporates actual smoke measurements, more comprehensive data sources, advanced modeling techniques, and a broader range of predictors to improve the accuracy and reliability of smoke and health forecasts.

Conclusion

This project aimed to answer critical questions about the impact of wildfire smoke on public health in Odessa, Texas, focusing on historical and forecasted trends in smoke density and health indicators. The analysis demonstrates that wildfire smoke density is projected to increase steadily from 2025 - 2050, a trend driven by historical patterns and modeled using the ARIMA approach. Additionally, health indicators—including respiratory conditions, cancer rates, and birth defects—are projected to worsen over the same period, showcasing the compounded impact of industrial emissions and wildfire smoke in the region.

The findings reveal that Odessa's public health is at significant risk, particularly given its location in the Permian Basin, an area already heavily burdened by industrial pollution. This study provides a foundation for understanding how these environmental factors intersect, emphasizing the need for actionable policies that address both industrial and wildfire-related pollution.

Through a human-centered approach, this study highlights the importance of transparency, inclusivity, and data literacy in interpreting and presenting results. By disclosing the limitations of the models and data, the project ensures stakeholders can make informed decisions without overreliance on uncertain projections. The focus on diverse health indicators ensures the findings are relevant to a broad spectrum of the population, particularly vulnerable groups.

To address these challenges, the study calls for the development of robust public health strategies within the next 5–10 years. These strategies should include enhanced healthcare access (respiratory and reproducibility), targeted wildfire prevention policies, and increased community awareness about the risks of smoke exposure. The findings serve as a guide for resilience and quality improvement of life in Odessa, offering critical insights for policymakers, healthcare providers, and residents as they navigate the challenges of a changing environmental and public health landscape.

References

Chevron Policy, G. and P. A. (2024, November 13). Permian basin. chevron.com. https://www.chevron.com/what-we-do/energy/oil-and-natural-gas/assets/permian#:~:text=As%20of%202021%2C%20the%20Permian,%3E1/3

Satellite data reveals extreme methane emissions from Permian Oil & gas operations; shows highest emissions ever measured from a major U.S. oil and Gas Basin. Environmental Defense Fund. (2020, April 22). https://www.edf.org/media/satellite-data-reveals-extreme-methane-emissions-permian-oil-gas-operations-shows-highest

ScienceDaily. (2024, July 16). Oil and natural gas development in Permian is a key source of ozone pollution in Carlsbad Caverns National Park, study finds. ScienceDaily. https://www.sciencedaily.com/releases/2024/07/240716122718.htm

Centers for Disease Control and Prevention. (n.d.). *How wildfire smoke affects your body*. Centers for Disease Control and Prevention. https://www.cdc.gov/wildfires/risk-factors/index.html

Centers for Disease Control and Prevention. (2023, August 18). *Outdoor workers exposed to wildfire smoke*. Centers for Disease Control and Prevention. https://www.cdc.gov/niosh/topics/firefighting/wffsmoke.html

Martha Pskowski, I. C. N., & Dylan Baddour, I. C. N. (2023, January 13). EPA moves away from Permian Basin Air Pollution Crackdown. The Texas Tribune. https://www.texastribune.org/2023/01/13/texas-permian-basin-air-polution-epa-oil/

Westerling, A. L., Hidalgo, H. G., Cayan, D. R., & Swetnam, T. W. (2006). Warming and earlier spring increase western U.S. forest wildfire activity. Science, 313(5789), 940–943. https://doi.org/10.1126/science.1128834

Dennison, P. E., Brewer, S. C., Arnold, J. D., & Moritz, M. A. (2014). Large wildfire trends in the Western United States, 1984–2011. Geophysical Research Letters, 41(8), 2928–2933. https://doi.org/10.1002/2014gl059576

Preisler, H. K., & Westerling, A. L. (2007). Statistical model for forecasting monthly large wildfire events in Western United States. Journal of Applied Meteorology and Climatology, 46(7), 1020–1030. https://doi.org/10.1175/jam2513

Jeon, S., & Han, S. (2023). Time series forecasting for air pollution in Seoul. arXiv. https://doi.org/10.48550/arXiv.2309.09579

Centers for Disease Control and Prevention. (n.d.-a). Compressed mortality, 1968-2016 request. Centers for Disease Control and Prevention. https://wonder.cdc.gov/cmf-icd8.html

Birth and death records. Texas Department of State Health Services (DSHS). (2024, October 1). https://www.dshs.texas.gov/

Admin. (2022, October 13). Inverse square law - statement, Formula and applications. BYJUS. https://byjus.com/physics/inverse-square-law/#:~:text=Inverse%20square%20law%20states%20that,the%20square%20of%20the%20distance%E2%80%9D.&text=%E2%80%9CThe%20intensity%20of%20the%20light,the%20observer%20to%20the%20source%E2%80%9D.

Ibm. (2024, August 19). What are Arima Models?. IBM. https://www.ibm.com/topics/arima-model#:~:text=What%20are%20ARIMA%20Models?,bigger%20than%20this%20weeks%20order.

Data Sources

Compressed Mortality 1968 - 1978, 1979 - 1998, 1999 - 2016

United States and Puerto Rico Cancer Statistics, 1999-2021 Incidence/Mortality

PLACES: County Data (asthma and chronic obstructive pulmonary disease), 2020 - 2024

Texas Birth Defects Registry (TBDR) Annual Report, 1999 - 2020