James Joko
DATA 512: HCDS
12/11/2023
Class Project Part 4: Final Report
Repository: https://github.com/jamesjoko/data-512-project
City: Santa Fe, NM

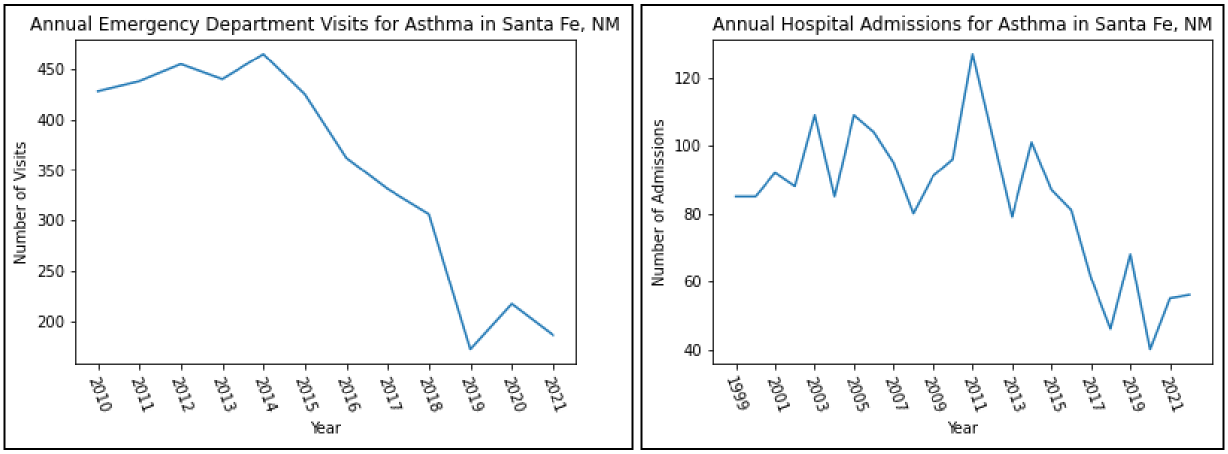
Introduction

For my DATA 512 course project, I set out to investigate and analyze the effects that wildfire smoke had on public health. This led me to look at the rate of high-risk asthma cases, a condition that is directly affected by poor air quality. Asthma is a chronic respiratory disease that involves the inflammation and narrowing of airways. People living with asthma can experience episodes and attacks that can vary from mild to life-threatening. They are especially at risk from wildfire smoke because particulates can agitate air pathways and exacerbate the symptoms or attacks of asthma. In times of poor air quality, patients with asthma experience an overall reduced quality of life and are advised to stay indoors and not partake in strenuous activities. According to the New Mexico Department of Health (NMDH)[2], 10.6 percent of New Mexico Adults had asthma in 2021, with the national median being 9.8 percent. Specifically to my assigned city, according to the American Lung Association[3], 9.8 percent of the total Santa Fe county population has asthma. Patients with asthma can lessen the severity of their condition by having routine healthcare visits, using proper medication, and self-management techniques. When implemented properly, this is referred to as “controlled asthma”. However, according to an epidemiology report[1], 52.8 percent of adults and 38 percent of children in 2007-2010 stated that they had poorly controlled asthma. Since a significant portion of the Santa Fe population suffers from asthma, I think it is imperative for decision-makers to both be aware of the prevalence of asthma in their community and take the necessary steps to alleviate the obstacles that people with asthma face in their quality of life and in times of poor air quality. I would also like to investigate if asthma affects different demographics disproportionately and inform policymakers on how to focus resources on specific groups of the community that especially need it. I think that this issue is very interesting from a practical standpoint. Since reducing the quantity and severity of wildfires is not entirely feasible, the problem is instead about easing access to proper asthma control by improving the accessibility of education/awareness, teaching and encouraging healthy behaviors for people with asthma, and improving clinical services outreach, especially in low-income communities.

Background/Related Work

My two hypotheses for the course project are as follows. First, there is a positive correlation between wildfire smoke exposure and asthma-related hospital admissions. Second, there is a positive correlation between wildfire smoke exposure and asthma-related emergency department visits. My intuition behind these hypotheses is that as air quality in Santa Fe worsens, people living with asthma are at a larger risk for severe asthma attacks or episodes. This would lead to an increase in admissions to hospitals or emergency departments.

The two datasets I used to extend my Part 1 results are asthma-related admissions to the hospital and visits to the emergency department for residents in Santa Fe County. The data comes from the New Mexico Environmental Public Health Tracking (NM EPHT). The NM EPHT operates within the Environmental Health Epidemiology Bureau (EHEB) of the New Mexico Department of Health Epidemiology and Response Division. This data is important because it tracks the number of asthma-related incidents per year, per county. I use this data to look for a correlation between my smoke estimate and ED/hospital admissions. An important detail to notice is a steep decrease in both metrics around 2015. This is explained further in the limitations section below.



Figures 1 and 2: Annual Asthma Emergency Department Visits (left) and Hospital Admissions (right)

The New Mexico Department of Health tracks data regarding asthma statistics among its population and also occasionally releases epidemiology reports. One especially helpful document was the NM DoH's Asthma Strategic Plan which outlines the current state of asthma and the state's ongoing outreach to mitigate obstacles that people with asthma face, such as poor routine clinical care in low-income communities. Other research has also been done to investigate the effects of wildfire smoke on New Mexico residents. First, a study performed by Resnick et al. examines the association between a particulate pollutant and ED visits for respiratory and cardiovascular conditions in Albuquerque, New Mexico after the Wallow fire of 2011[6]. They found a positive association between the two variables, suggesting that wildfire smoke does increase asthma-related emergency incidents. Second, a study conducted by Stowell et al. used dynamically downscaled climate model projections to estimate additional asthma ED visits and hospitalizations due to a particulate pollutant in the Western US in the 2050s[7]. The study domain includes data from New Mexico. The researchers found that the results suggest substantial increases in health burdens from exposure to wildfire smoke may occur by 2050. Specifically for New Mexico, an estimated 1359 excess fire-related asthma events would occur in the 2050s. The conclusions and results of the research support my initial hypothesis. Since I could not find research that looked specifically at smoke or asthma-related events in Santa Fe County, I did not have any external material to inform my analysis or system design.

During Project Part 2, I considered developing a forecasting model to predict asthma-related ED/hospital admissions using the smoke estimate. I ultimately decided to not proceed with a model because I did not feel I had sufficient data to make a robust model and feared that the implications would lead stakeholders in an incorrect direction. This is covered in depth in the discussion/implications section below.

Methodology

My smoke estimate metric for an individual fire is calculated using three variables: acres burned by the fire, distance from Santa Fe to the closest point on the fire polygon, and whether the fire was prescribed or not. Once the smoke estimate was calculated for each fire, I summed the metric within each year to obtain an annual estimate. The reasoning for these three factors is as follows. I am assuming a larger fire with more acres burned would produce more smoke than a smaller fire. I divide this number by the distance from the closest point of the fire to Santa Fe. A fire farther away from Santa Fe would put less smoke in it than a similar-intensity fire close by. Lastly, a prescribed fire will produce significantly less smoke than a wildfire. Therefore, I scaled the smoke estimate down by a third when the source was a prescribed fire. Lastly, I added a factor of 250 in the denominator to align it closer with AQI. I then compared the estimate with AQI to find a similar increasing trend in smoke estimate and AQI. Figure 4 below shows that the AQI and smoke estimate in Santa Fe have been consistently increasing since 1980. The smoke also has high fluctuation but is generally showing an increase.

$$\boxed{SmokeEstimate(Fire) = \begin{cases} \dfrac{(FireSize)}{(FireDistanceToSF * 750)} & \text{if Fire is not Prescribed} \\ \dfrac{(FireSize)}{(FireDistanceToSF * 250)} & \text{if Fire is Prescribed} \end{cases}}$$

Figure 3: Smoke Estimate Calculation

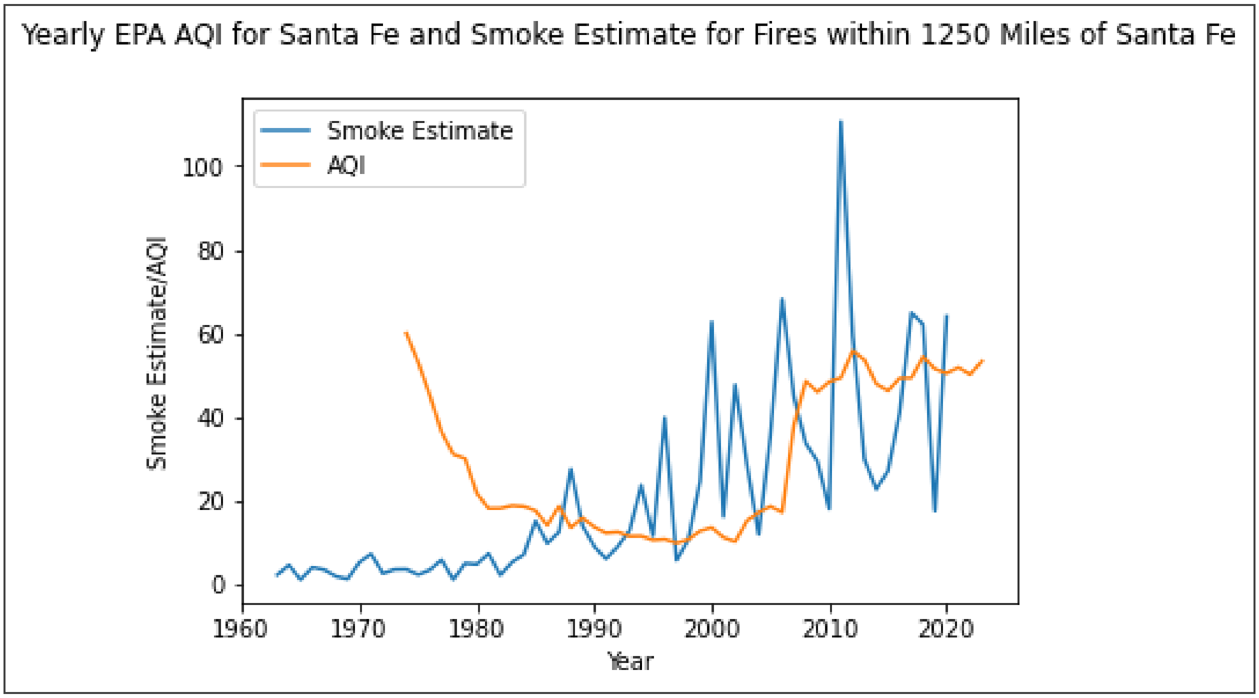


Figure 4: Smoke Estimate vs AQI

To calculate the correlation between the calculated smoke estimate and asthma-related ED/hospital visits, I used Spearman's correlation. Spearman's correlation has three assumptions. First, the compared variables are measured on an ordinal, interval, or ratio scale. Both variables are measured on a ratio scale. Second, the variables represent paired observations. The variables are paired because they are annual observations. Third, there is a monotonic relationship between the two variables. Since the smoke estimate generally increases and the ED/hospital visits generally decrease, there is a monotonic relationship between the two variables. When considering statistical tests, I also considered alternatives. One alternative was the Pearson correlation. My analysis did not use this test because one of its assumptions is that data from both variables follow normal distributions. I do not believe both variables satisfy this assumption. In addition, I do not believe both variables are stationary, as I expect their statistical properties to change over time. The second alternative was a Kendall's Tau correlation. My analysis does not use this test because it is used for ordinal data, while my variables are continuous/ratio.

A human-centered consideration I made was to include the utmost accuracy in my analysis due to the importance of the potential findings. After looking deeper into why there was a sudden dip in asthma-related ED/hospital admissions, I found that an updated version of ICD coding was implemented in 2015, which led to a redefinition of hospital admission and emergency department visits. Due to the limitation of the asthma-related data below, I separated the datasets into two different time frames to account for the differing ICD versions. I analyzed both time frames and recorded my findings.

Findings

ICD9 Coding

When performing a Spearman's correlation test on smoke estimate and asthma ED visits for IDC9 coding, the correlation coefficient was 0.14285714285714288 with a p-value of 0.7871720116618075. This suggests that the two variables are positively correlated slightly, meaning as smoke increases, asthma ED visits also increase. This can visually be seen in Figure 5 below. However, the p-value generated by the test is very high, indicating that there was insufficient evidence to reject the null hypothesis that there is no correlation between the two variables. Therefore, the test was not statistically significant. When performing a Spearman's correlation test on the smoke estimate and asthma hospitalizations for IDC9 coding, the correlation coefficient was 0.3 with a p-value of 0.6238376647810728. This suggests that the two variables are positively correlated, with a stronger relationship than the previous test. This means as smoke increases, asthma hospitalizations also increase. This can visually be seen in Figure 6 below. The implications of a high p-value mentioned above also apply here. Visually, this also looks like a strong correlation because peaks and valleys generally match.

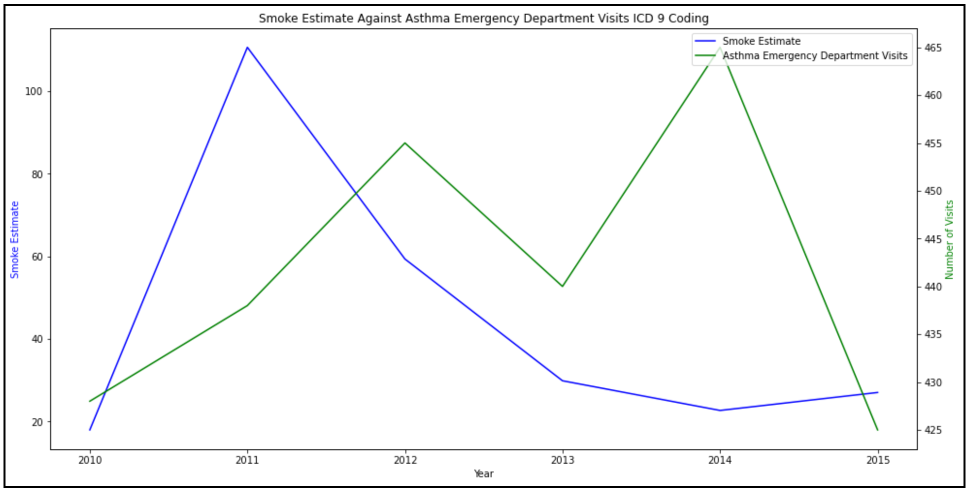


Figure 5: Smoke Estimate against Asthma ED Visits under ICD9 Coding

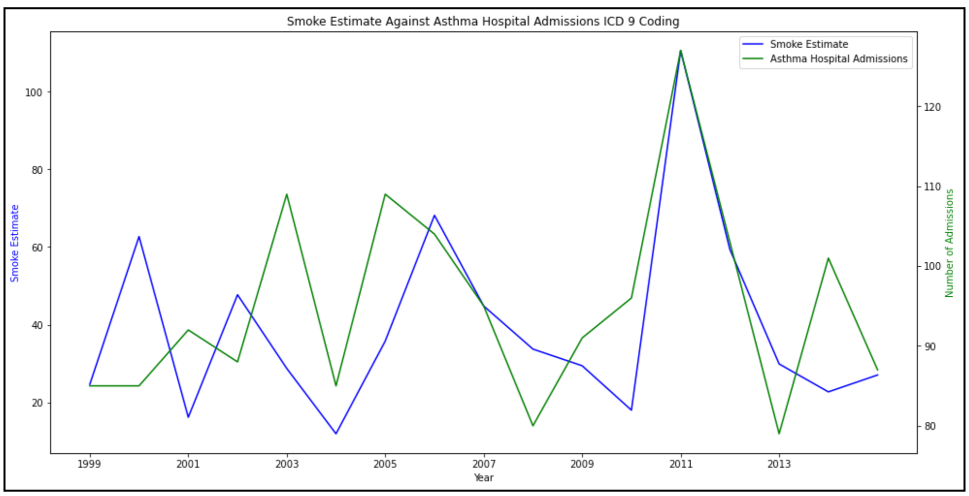


Figure 6: Smoke Estimate against Asthma Hospital Admissions under ICD9 Coding

ICD10 Coding

When performing a Spearman's correlation test on the smoke estimate and asthma ED visits for IDC10 coding, the correlation coefficient was 0.2950229051153235 with a p-value of 0.2503193417519162. This suggests that the two variables are somewhat positively correlated, meaning as smoke increases, asthma ED visits also increase. This can visually be seen in Figure 7 below. The p-value is significantly smaller on this test than the two tests performed for the ICD 9 coding, but the value is still not small enough to be statistically significant. When performing a Spearman's correlation test on the smoke estimate and asthma hospitalizations for IDC9 coding, the correlation coefficient was -0.6 with a p-value of 0.28475697986529375. This suggests that the two variables have a moderately strong negative correlation. This means as smoke increases, asthma hospitalizations decrease and vice versa. This can visually be seen in Figure 8 below. Once again, the p-value is small, suggesting some evidence to reject the null hypothesis, but not small enough to be statistically significant.

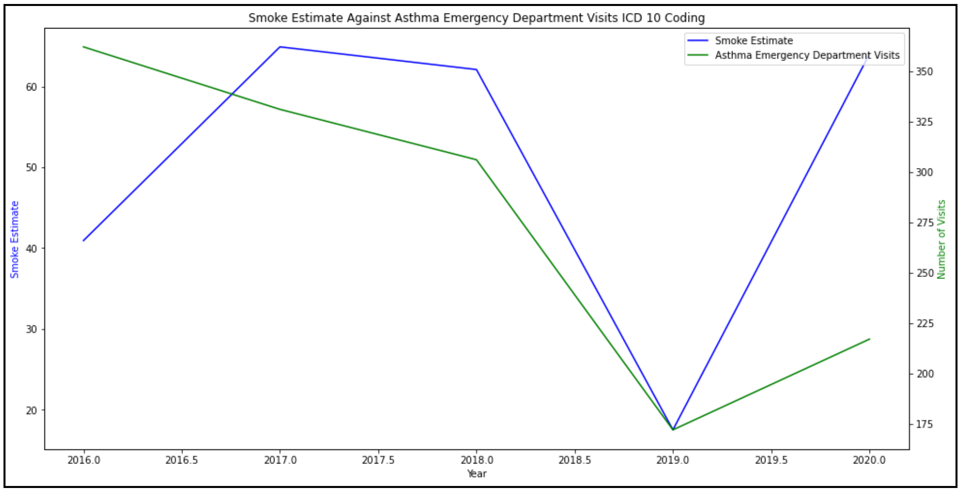


Figure 7: Smoke Estimate against Asthma ED Visits under ICD10 Coding

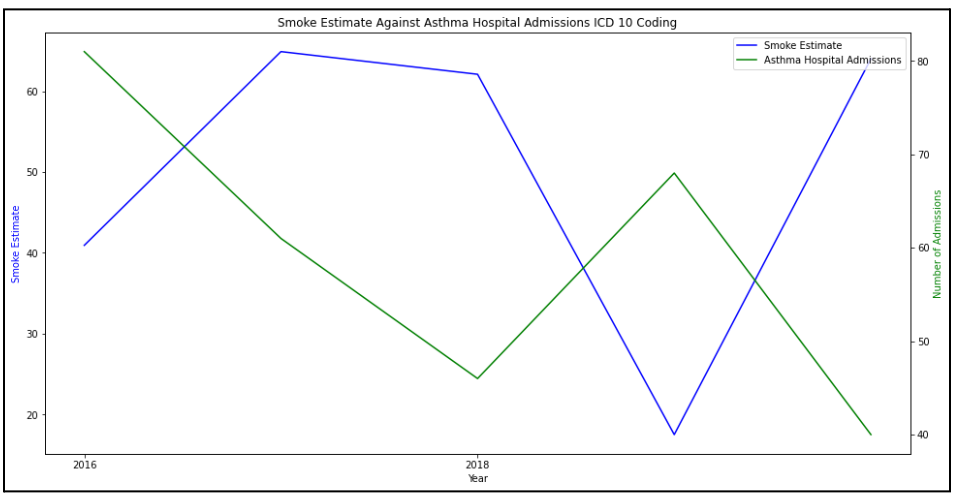
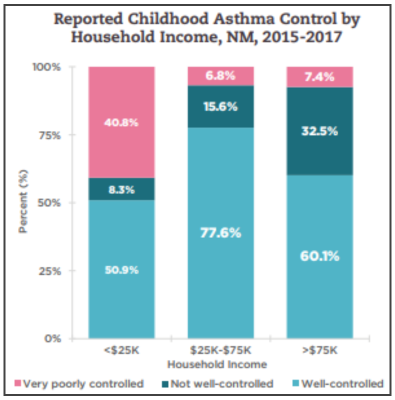


Figure 8: Smoke Estimate against Asthma Hospital Admissions under ICD10 Coding

Discussion/Implications

My findings suggest three takeaways: First, I found a slight positive correlation between the smoke estimate and asthma hospitalizations/ED visits pre-2016. Second, I found a slight positive correlation between the smoke estimate and asthma ED visits post-2016. Lastly, I found a moderate negative correlation between the smoke estimate and asthma hospitalizations post-2016. However, the statistical significance of these findings is not strong enough to generate any strong implications due to the lack of data. One takeaway that should be generated from this analysis is to increase the frequency of reporting on asthma-related metrics monthly to generate more data and capture more nuance for specific fire events and seasons. Secondly, while there were no statistically significant results of this study, there are still action steps that policymakers and stakeholders should employ to improve the quality of life for people living with asthma in Santa Fe. For example, in New Mexico, asthma disproportionately affects children more than adults and lower-income households struggle more with properly managing their asthma severity[4]. See Figures 9 and 10 below. Therefore, two action steps for policymakers are as follows. First is to further educate the population of asthma triggers, especially by giving guidance to children in schools and informing households where smoking is prevalent. The second action step is to provide health services such as free popup clinics in lower-income communities to allow for easier access to routine healthcare visits. Since one in every ten Santa Fe residents is currently diagnosed with asthma, I believe it is imperative that policymakers act as soon as possible. Patients with asthma can lessen the severity of their symptoms and risk of episodes/attacks when given education on self-management techniques and easier access to routine healthcare.



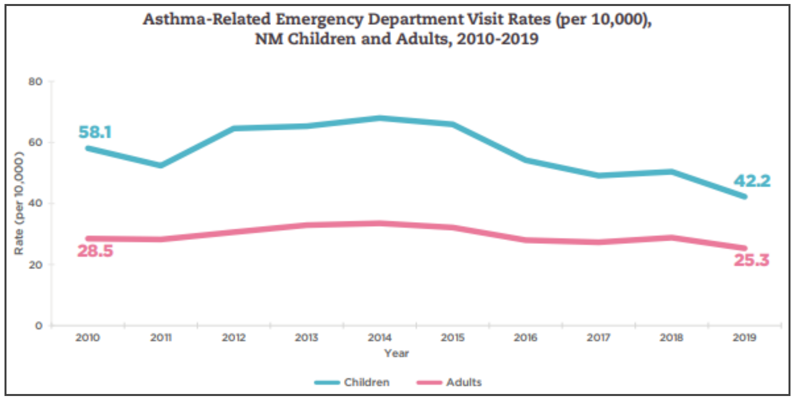


Figure 9 (left): Childhood Asthma Control by Household Income.

Figure 10 (right): Asthma-Related Emergency Department Visit Rates (per 10,000)

Reflection

One aspect of my analysis is that it does not include any statistical models. This was a conscious decision because of the limitations of my data, which are explained in the limitations section below. Since I only have a very limited amount of data with the old and new ICD coding, I believe I have insufficient data with either time frame to adequately make a forecasting or predictive model. I do not want to mislead the viewer on what implications come from the analysis, so I keep my analysis looking for statistical correlation instead to see whether smoke affects asthma-related ED/hospital visits.

Homework 1 and the class readings on reproducibility were essential in defining the reproducibility standards for my project. When we evaluated Hickey's analysis of a film's strength of female characters on financial performance in the Week 2 reading reflection, we were able to identify poor practices reproducibility. On the other hand, we could also identify the best reproducibility practices when analyzing Dr. Keegan's redone analysis. Throughout all stages of my project, I made my utmost effort to ensure that any viewer could read my notebooks and understand the thought process behind actions taken within them. I also made it very easy to replicate my results as I defined which order to run notebooks in and each notebook can be run from top to bottom to reproduce the same results. In fact, this was very helpful for me at the start of project part 3 when I accidentally deleted all of my intermediate data. When looking back at my notebook to create the intermediate data files, I simply had to rerun the notebook from top to bottom to regenerate my intermediate data files and look at specific comments to understand which cell does what. Another way I employed HCDS principles is by carefully choosing and vetting my sources. A large majority of my learning and sources for this project were obtained directly from the local government's public releases such as epidemiology reports and data sources. I also did my best to mitigate any bias by using observational language in my report and repository and fully explaining any conclusions I came to.

Limitations

Smoke Estimate Accuracy

The biggest limitation of the project is the accuracy of the smoke estimate. My current smoke estimate calculation uses three factors to estimate the impact of smoke on Santa Fe: the number of acres burned by the fire, the distance from Santa Fe, NM to the closest point on the fire polygon, and whether the fire is classified as prescribed or wildfire. While the calculation does contain major factors to smoke's effect on Santa Fe residents, it does not tell the full picture when considering the effect of a specific fire's smoke. This can potentially lead to incorrect findings and implications. For example, the assumption that prescribed fires emit one-third of the amount of smoke of a similar wildfire may not apply to every prescribed fire. We do not have the available data to properly measure what the exact fraction is, but we do know that prescribed fires generally emit less smoke than wildfires. Additionally, other data fields are missing to calculate a more accurate smoke estimate such as the wind direction and magnitude during the fire's duration, the amount of flammable material in the impacted area, and the topography of the impacted area. For example, a large wildfire could have occurred a moderate distance from Santa Fe, which would add a significant amount to the annual smoke estimate, but the factors listed previously could have prevented any smoke from reaching Santa Fe.

The dataset used to calculate the smoke estimate also contains many flaws that can lead to an inaccurate smoke estimate. First, the dataset merges many existing wildfire and prescribed fire datasets that are limited in some way either through poor data collection, specific spatial footprints, or lack of fire metadata. Secondly, since wildfire mapping before 1984 was inconsistent, infrequent, and done without the aid of more modern fire mapping methods, areas burned before 1984 underestimate the actual land mass burned. Thirdly, prescribed fire data in the dataset represents only a fraction of the area burned in prescribed burns across all years due to lack of reporting, particularly on private lands. This error is more pronounced in the earlier years of the dataset, particularly in the southeastern U.S.

Timeline of Datasets

One non-ideal aspect of the data is that the earliest annual record is 1999 for hospital admissions and 2010 for emergency department visits for Santa Fe residents. The data for my smoke estimate in Santa Fe dates back to 1963. Therefore, I was only able to incorporate asthma hospital admissions into my smoke estimate for 21 years and incorporate asthma hospital admissions for 12 years. Furthermore, I am only able to analyze the time frames from pre-2016 and post-2016 due to a change in ICD coding, a limitation listed below. Therefore, I am only able to individually assess 1999-2015 and 2016-2022 for asthma hospital admissions and 2010-2015 and 2016-2021 for emergency department visits. Due to the low quantity of data, my findings may have insufficient evidence to be fully reliable. To perform a more robust analysis, more asthma-related data with consistent ICD coding is required to properly analyze the effects of smoke on severe asthma-related incidents. See Figure 11 below where green indicates ICD10.

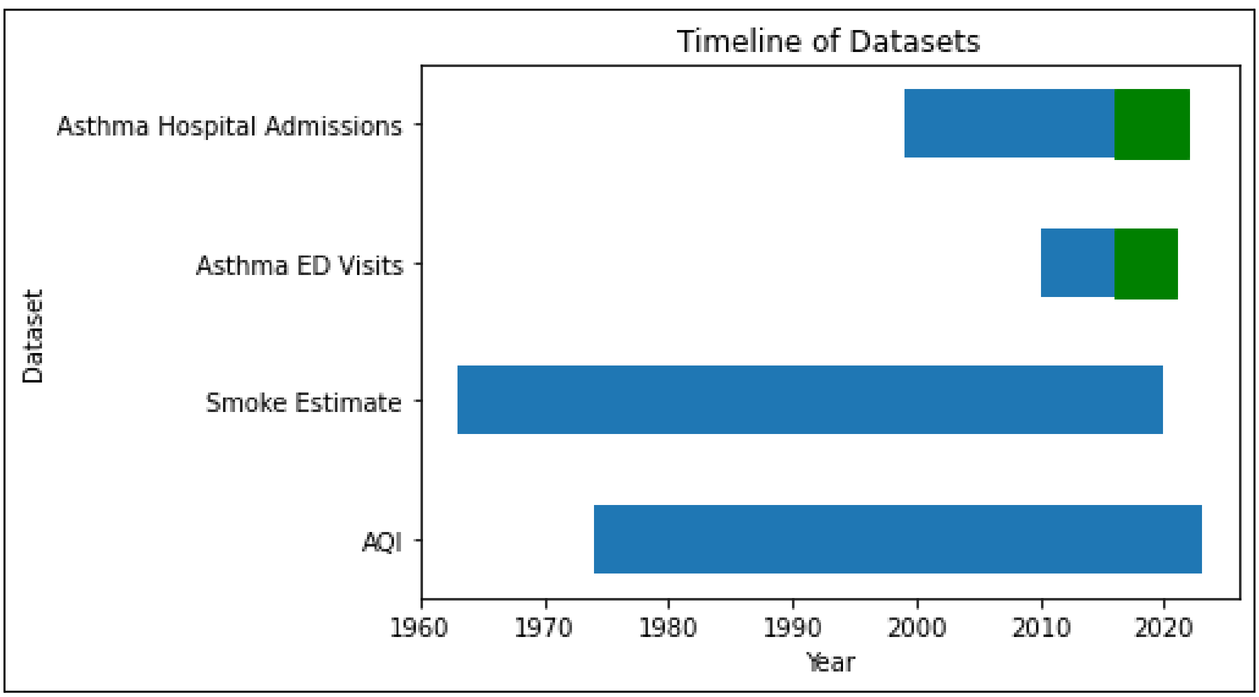


Figure 11: Timelines of Project Datasets

Incomplete Asthma Data

In conjunction with the limitation above, I reached out to the NM EPHT via email on 11/17 asking for access to the raw dataset instead of having to manually query individual records of information from the website. Having to manually query and record each specific piece of information (such as asthma ED visits in 2000 for the 0-14 age group, Hispanic ethnicity, etc.) would be incredibly tedious and time-consuming. This led me to only perform analysis on total annual admission counts instead of looking further into age, race, and sex demographics. I received a response on 12/6 reading "Access to the raw dataset would require an Internal Review Board (IRB) approval from your institution as NMDOH cannot release protected health information without it". The approval process would likely take longer than the remaining course duration, leading me to continue with the analysis using only total annual counts for asthma hospital admissions and ED counts.

Sudden Dip in Asthma-Related ED/Hospital Admissions

As shown above, in the findings section, asthma-related emergency department visits suddenly decreased around 2015. Additionally, asthma-related hospitalizations suddenly decreased around 2016. The reason for this is that the coding system called the International Classification of Diseases (ICD) underwent a major revision that went into effect on October 1st, 2015. The ICD is used to classify diagnoses for hospital and emergency department visits. Therefore, the two versions do not provide comparable results and the two time periods should be analyzed separately. The prevalence of currently diagnosed asthma stays relatively consistent over the full timeline for both adults and children. This further impacts the timeline of dataset limitation since my analysis cannot combine the two time periods.

Conclusion

My project set out to find a statistically significant correlation between an estimated smoke impact on Santa Fe residents and asthma-related ED/hospital admissions. While I was unable to find statistically significant findings due to a lack of quality data, there are still action steps that policymakers can take to improve the quality of life for people living with asthma in Santa Fe. These steps include educating the population on asthma triggers, especially in schools, and improving access to routine healthcare visits, especially in low-income communities.

This study should inform your understanding of HCDS by showing how important reproducibility is for other researchers to build off of previous analysis, closely vetting data sources and techniques to prevent bias or misguided implications, and mitigating our own confirmation bias when drafting a hypothesis and performing analysis that may or may not support our initial attitudes. The ultimate purpose of HCDS is to be transparent and honest in the analysis we perform because the implications of our work can have real-world ramifications and impacts.

References

1. The Burden of Asthma in New Mexico: 2014 Epidemiology Report. New Mexico Department of Health. Santa Fe, NM. January 2014. https://www.nmhealth.org/data/view/environment/54/.
2. Lashway, S., Whiteside C., Pell, D., Health Behaviors and Conditions of Adult New Mexicans. New Mexico Department of Health. Santa Fe, NM. 2021. https://www.nmhealth.org/data/view/behavior/2826/
3. "New Mexico | State of the Air." *American Lung Association*, www.lung.org/research/sota/city-rankings/states/new-mexico/santa-fe. Accessed 11 Dec. 2023.
4. Clearing the Air Reducing the Burden of Asthma in New Mexico. New Mexico Department of Health. Santa Fe, NM. 2021. https://www.nmhealth.org/publication/view/plan/6655/
5. Wilkerson, D. The Burden of Asthma: Hospitalization and Emergency Department Visits in New Mexico, 2016. New Mexico Department of Health. Santa Fe, NM. 2017. https://www.nmhealth.org/data/view/report/2076/

6. Resnick, Adam, et al. "Health outcomes associated with smoke exposure in Albuquerque, New Mexico, during the 2011 Wallow fire." *Journal of public health management and practice* 21 (2015): S55-S61. https://www.jstor.org/stable/48517478
7. Stowell, Jennifer D., et al. "Asthma exacerbation due to climate change-induced wildfire smoke in the Western US." *Environmental Research Letters* 17.1 (2022): 014023. https://doi.org/10.1088/1748-9326/ac4138

Data Sources

For information regarding the use, citation, terms of use, data structure, data fields, and issues or special considerations, please see the README file in my project repository. Since the handout only asks for links, I will not include additional information here.

1. Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons)
2. Air Quality System (AQS) API
3. New Mexico EPHT Primary Asthma ED Visit Counts
4. New Mexico EPHT Asthma Hospital Admissions