ENVIRON 710

Final Project Report

Spring 2026

Impacts of Wildfire Activity on Air Quality and Social Vulnerability in California:

A County-level Analysis

By

Abrao Soares Pereira

Cecilia Liang

Flannery McNair

Introduction and Background

California has seen a significant increase in wildfires over the past twenty years, with more frequent and severe fires spreading across larger areas. This is due to rising temperatures, long periods of drought, and the growth of human settlements near wildlands (Anthony et al., 2011). The amount of land burned each year in the state has increased about eight times since 1970, and recent fire seasons have repeatedly broken previous records (IQAir, 2023). Wildfire smoke, which is a complex mix of fine particles (PM2.5), nitrogen oxides, and volatile organic compounds, can significantly reduce air quality in large areas for days or weeks. The effects of this smoke go far beyond the area directly affected by the fire (Hadiwijoyo et al., 2025). The January 2025 Los Angeles wildfires illustrated this vividly: daily average PM2.5 concentrations exceeded 100 μg/m³ at downtown monitoring stations during peak smoke days (Schollaert et al., 2025), pushing an entire metropolitan population into hazardous exposure conditions. These events make clear that wildfire activity has long contributed to air quality degradation, but its impact has become increasingly significant in recent years.

The public health consequences are substantial. Short-term exposure to wildfire-derived PM2.5 is associated with increased emergency department visits for respiratory illness, worsening asthma, and cardiovascular events, while long-term exposure is linked to reduced lung function, chronic disease progression, and premature mortality (Reid et al., 2016; Tillakaratne et al., 2023). The 2020 fire season, which set a new California record for area burned, drove PM2.5 and ozone to hazardous levels across the Bay Area and Central Valley; modeled health impacts from that season alone were estimated to include thousands of premature deaths and hundreds of thousands of respiratory illness days (Carreras-Sospedra et al., 2024). Despite this growing body of evidence, most studies have focused on individual fire events or short temporal windows. We lack a comprehensive understanding of how total annual wildfire-burned area relates to Air Quality Index (AQI) trends at the county level across California over a multi-year period, particularly when controlling for co-occurring drivers such as temperature, drought severity, and urbanization.

Crucially, the risks from wildfire smoke are not distributed equally. Socioeconomic and demographic characteristics shape both the likelihood of exposure and the capacity to respond to environmental hazards. Research using the CDC Social Vulnerability Index (SVI) has shown that communities characterized by minority status, limited English proficiency, lower educational attainment, and crowded housing experienced the largest increases in heavy smoke-exposure days between 2011 and 2021 (Vargo et al., 2023). For example, individuals in lower-income or minority communities are more likely to work in outdoor or high-exposure occupations, have limited access to protective resources such as air filtration, and face barriers to receiving timely public health information due to language constraints. Housing conditions, including crowding and lower-quality infrastructure, may further reduce

the ability to mitigate indoor smoke exposure, while limited access to healthcare and transportation can constrain adaptive responses during extreme events. At broader spatial scales, recent studies highlight that the impacts of wildfire smoke are also unevenly distributed across regions and populations. For instance, Do et al. (2024) document substantial spatial heterogeneity in the respiratory health impacts of wildfire-derived $PM_{2.5}$ across California, suggesting that identical exposure levels can produce different health outcomes depending on local vulnerability factors. Similarly, Cleland et al. (2025) show that wildfire smoke exposure often co-occurs with extreme heat events, compounding risks in already vulnerable communities and exacerbating existing inequities in environmental burden. Additional research has emphasized that highly vulnerable populations, including individuals experiencing homelessness, face disproportionate exposure due to limited shelter and reduced adaptive capacity (Zheng et al., 2025). Consistent with this literature, county-level analyses indicate that higher fire-derived $PM_{2.5}$ concentrations are associated with lower adaptive capacity and higher vulnerability scores and that highly vulnerable counties experience disproportionately greater increases in annual unhealthy air quality days (Jung et al., 2024).

Taken together, these findings suggest that the relationship between wildfire activity and air quality degradation is embedded within broader systems of social and environmental inequality. This raises the possibility that the statistical association between wildfire activity and AQI may itself vary across levels of social vulnerability, an equity-focused question that has received limited empirical attention and motivates the present analysis.

We address these gaps using county-level data across California from 2010 to 2025. We pursue three research questions. First, is the total wildfire burned area associated with the annual mean AQI at the county level after controlling for temperature, drought conditions, and percent urban land cover? We hypothesize that larger burned areas will be positively associated with higher mean AQI, even after accounting for these confounders. Second, in Los Angeles County specifically, is the burned area associated with an increase in the number of days with an AQI above 100—the threshold marking conditions unhealthy for sensitive groups? We hypothesize that years with greater burned area in or near the county will show significantly more unhealthy air quality days. Third, we examine whether the association between wildfire activity and AQI varies by county-level social vulnerability. While wildfire smoke is driven primarily by environmental and atmospheric processes, prior research suggests that exposure patterns and baseline environmental conditions differ across communities. Therefore, we assess whether the relationship between burned area and AQI differs across levels of social vulnerability, without imposing a strong directional hypothesis. Identifying such differences would have important implications for environmental management, targeted public health interventions, and equitable emergency preparedness planning.

Methods

To find answers for our three research questions and the hypothesis we fit 3 models. For RQ1, we used a linear mixed-effect model with county-level random intercepts to examine the association between log-transformed burned area and annual mean AQI across 53 counties in California. For RQ2, we used Poisson generalized linear model with monthly observations from Los Angeles County to examine the relationship between burned area and counts of days with AQI above 100. Finally for RQ3, we extended the RQ1 model by adding an interaction term between burned area and social vulnerability index.

Wildfire Data:

Wildfire perimeter data were obtained from the California Department of Forestry and Fire Protection Historical Fire Perimeters dataset. The dataset includes ignition date, burned area (acres), and fire name for fires meeting minimum size criteria or those causing significant damage. Using the sf package in R, we spatially joined fire perimeters to California county boundaries and aggregated records to the county-year level for 2010–2025. For each county-year, we calculated total burned area (acres) and fire event count as our primary wildfire activity measures. Because the burned area was strongly right-skewed, we applied a log~10~ transformation prior to inclusion in all regression models. County-years with no recorded fires were assigned a burned area of zero; to accommodate the log transformation, these were set to log~10~(1) = 0, indicating no fire activity.

Air quality:

Daily Air Quality Index (AQI) data were obtained from the U.S. Environmental Protection Agency's Air Quality System (AQS). For each county and year, we derived two outcome variables: (1) annual mean AQI and (2) the annual count of days with AQI > 100, corresponding to conditions classified as "Unhealthy for Sensitive Groups" or worse. Where multiple monitoring stations existed within a county, we averaged daily values across stations before computing annual summaries. County-years with fewer than 200 valid monitoring days were excluded from analysis to avoid unreliable annual averages, and we note that monitoring coverage is uneven across California's 58 counties—more rural counties in the Central Valley and northern Sierra Nevada have sparser networks—which may introduce measurement error that biases associations toward the null in less monitored areas.

Social Vulnerability Index (SVI):

Social vulnerability was measured using the CDC/ATSDR Social Vulnerability Index (SVI). We obtained county-level SVI data for 2022, including the four theme-specific percentile rankings and the overall composite ranking. Because percentile rankings are not directly comparable across years due to changes in census tract boundaries, we used the raw theme-specific percentages as time-varying covariates in regression models and also constructed a standardized overall vulnerability score by averaging the four theme ranks for each county-year, following CDC guidance.

Statistical Analysis:

All analyses were conducted in R. Prior to modeling, we screened all continuous independent variables for multicollinearity using variance inflation factors (VIF); any variable with VIF > 5 was examined for removal or consolidation with correlated predictors. We also inspected pairwise correlation matrices and histograms of all predictors.

To address the association between wildfire-burned area and air quality, we employed Linear Mixed-Effects Models (LMM) to account for the non-independent structure of our data. Since our dataset includes repeated annual observations across multiple California counties, a standard linear regression would violate the assumption of independent sampling units. By utilizing county-specific random intercepts, we allowed each county to have its own baseline Air Quality Index (AQI) while estimating the fixed effects of wildfire predictors across the state. Following the principles of ordinary least squares, we log-transformed the "Total Acres Burned" to address data skewness and mean-centered the Social Vulnerability Index (SVI) to ensure the interpretability of our interaction terms.

Our primary analysis examined the relationship between annual burned area and mean AQI, while an extended model included an interaction term between burned area and standardized SVI to test for effect modification in vulnerable populations. For the single-county analysis of Los Angeles, we utilized simple linear regression with a time trend, as grouping factors are not applicable to a single entity. To validate our models, we evaluated QQ plots to confirm the normality of residuals and residual-versus-fitted value plots to assess homoscedasticity. We also utilized Akaike Information Criteria (AIC) to ensure our model selection was parsimonious, avoiding overparameterization while maintaining high predictive accuracy. Throughout, we emphasize the interpretation of effect sizes, specifically, the change in mean AQI, or unhealthy day count, associated with a given change in burned area rather than focusing solely on statistical significance.

Results

The dataset for this analysis included 188 county-year observations spanning 53 California counties across four years: 2016, 2018, 2020, and 2022. Annual mean AQI ranged from approximately 25 to 125 across county-years, with a statewide mean of roughly 57 AQI points. Total acres burned varied dramatically across county-years, from near zero to over 4 million acres in 2020, reflecting the extreme right skew that motivated log10 transformation prior to modeling. This dataset was primarily used for research questions 1 and 3, as they were doing analyses across all counties in California for all four years. Monthly AQI and burned area data was used for research question 2, since it was only looking at Los Angeles county, which allowed the analysis to be more granular temporally.

Research Question 1: Is the total wildfire burned area associated with the annual mean AQI at the county level?

Prior to model fitting, all continuous predictors were screened for multicollinearity using variance inflation factors (VIF). All three predictors, log-transformed burned area, social vulnerability index, and year returned VIF values below 1.05 (log_acres = 1.029, SVI = 1.025, year = 1.005), well below the threshold of 5, indicating no meaningful multicollinearity among predictors.

Three candidate models were compared using Akaike Information Criterion (AIC): a standard generalized least squares regression without random effects (AIC = 1616.1), a base linear mixed effects model with burned area and year as fixed effects and a county-level random intercept (AIC = 1444.0), and an extended model additionally including social vulnerability index as a covariate (AIC = 1440.7). The mixed effects model with SVI was selected as the best-fitting model, representing a reduction of 175.4 AIC points over the standard regression and 3.3 AIC points over the base mixed model. The substantial improvement over the standard regression confirms that accounting for county-level baseline differences in AQI was essential in counties differing from one another in baseline air quality by approximately 15 AQI points on average, as reflected in the county-level random intercept standard deviation of 14.9.

Model diagnostics supported the validity of the selected model. Visual inspection of the residuals versus fitted values plot revealed no systematic pattern or funnel shape, indicating that the homoscedasticity assumption was reasonably met. The QQ plot of standardized residuals showed close adherence to the diagonal reference line through the central range of the distribution, with minor departures in the tails consistent with slightly heavy-tailed behavior. These departures were considered acceptable given the sample size of 188 observations and did not indicate a fundamental violation of model assumptions.

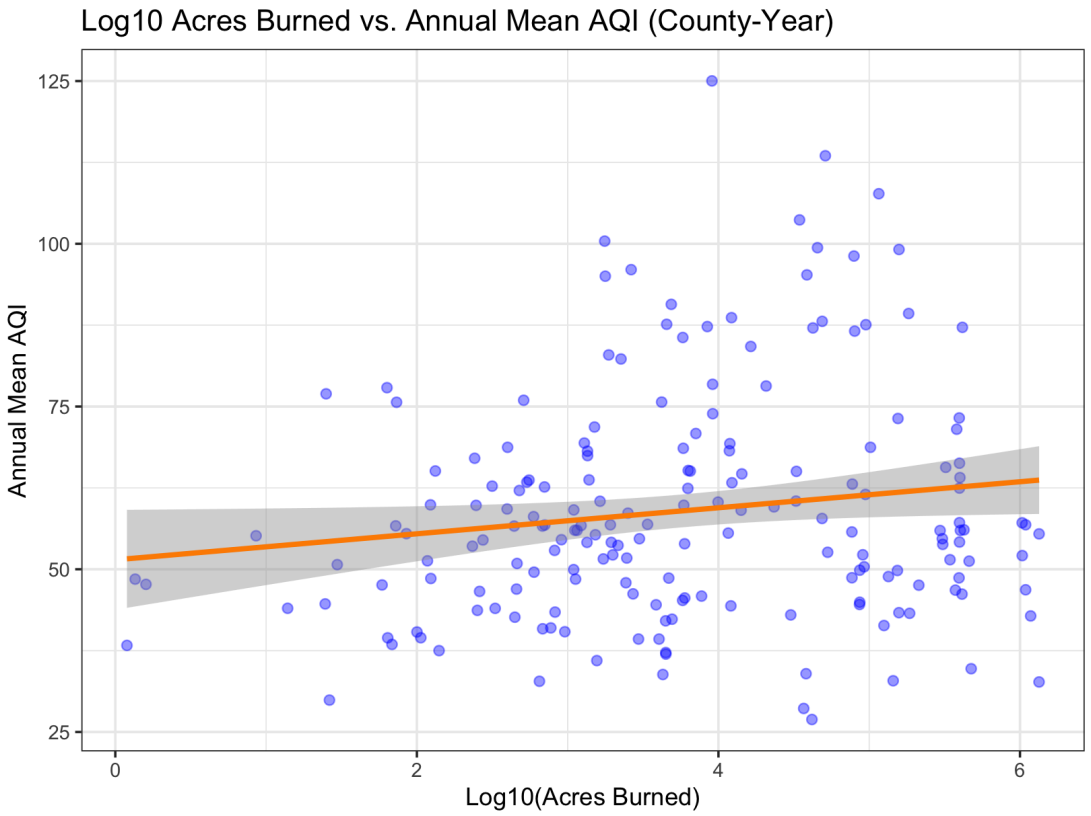


Figure 1: Model diagnostic plot

The selected model revealed that log-transformed wildfire burned area was a statistically significant positive predictor of annual mean AQI (β = 1.45, 95% CI [0.47, 2.43], p = 0.005). A one-unit increase in log10-transformed burned area corresponding to a 10-fold increase in raw acres that was associated with a 1.45-point increase in annual mean AQI after controlling for social vulnerability and year. Across the full observed range of wildfire activity in our dataset, this translates to an estimated 10-point difference in predicted AQI between counties with minimal fire activity and those experiencing the most extreme fire years. Social vulnerability index was also a statistically significant predictor of annual mean AQI (β = 14.40, 95% CI [2.4, 26.0], p = 0.020), indicating that counties with maximum social vulnerability scores were associated with mean AQI approximately 14.4 points higher than the least vulnerable counties, independent of wildfire activity. Year was not a statistically significant predictor (β = -0.29, 95% CI [-0.80, 0.22], p = 0.267), suggesting no detectable secular trend in AQI across the four study years after accounting for fire activity and social vulnerability. While wildfire activity explains variation in AQI across counties, these effects may differ across populations, motivating our examination of social vulnerability in RQ3.

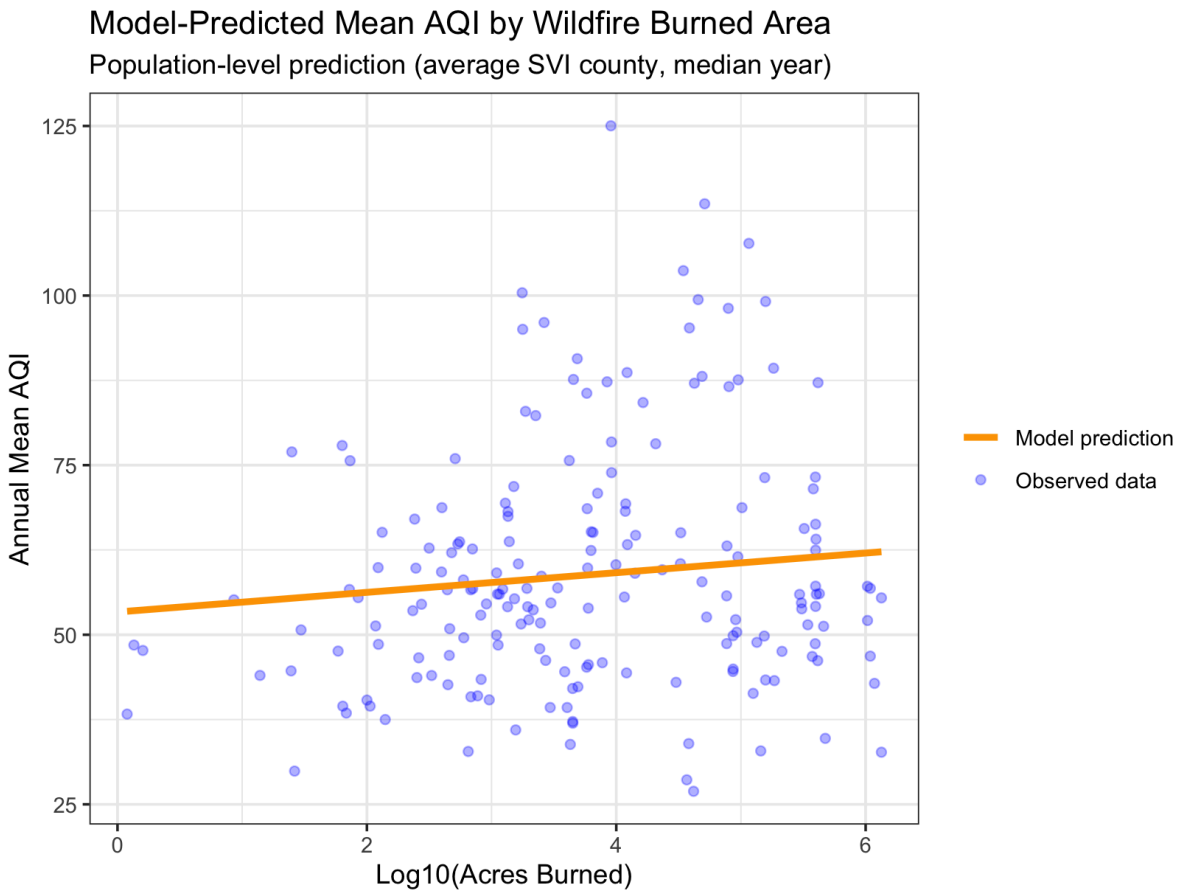


Figure 2: Fitted best model

Research Question 2: In Los Angeles County, is the burned area associated with an increase in the number of days with an AQI above 100?

The analysis for our second research question used monthly averages of AQI and burned area to analyze the relationship between these variables in Los Angeles county. The mean AQI each month across the four years we looked at was 90, with a standard deviation of 25. Additionally, the mean number of days each month where the AQI was over 100 was 9.2, with a standard deviation of 8.6.

A poisson regression was used to analyze the relationship between days with an AQI over 100 and burned area. The results indicated a statistically significant positive association between acres burned and the number of unhealthy air quality days ($p < 0.001$). The model showed that for each log-unit increase in burned acres, there is an estimated 40% increase in expected number of days with AQI > 100 ($p < 0.001$). Model fit improved compared to the null model (residual deviance = 240.73 vs null deviance = 381.93), suggesting that wildfire activity explained a meaningful portion of the variation in air quality outcomes.

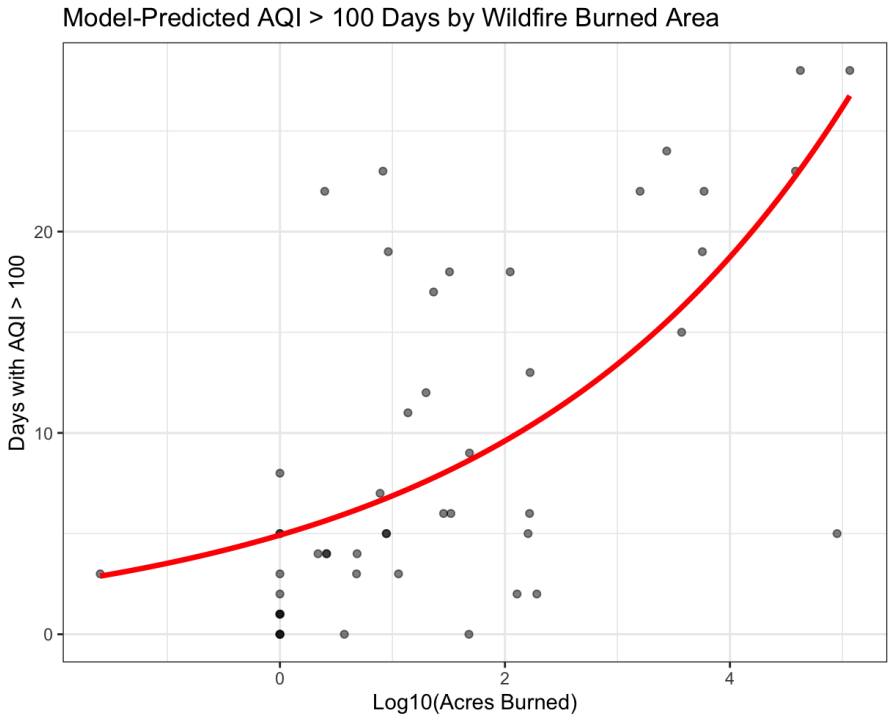


Figure 3: Model visualization

Figure 3 visualizes the relationship between the log burned area and number of days with an AQI > 100, along with the fitted values from the Poisson regression model. The plot shows a clear upward trend, consistent with our results from the regression model. The relationship is non linear, as the number of unhealthy AQI days increased more rapidly at higher levels of burned area.

Research Question 3: Does Social Vulnerability Modify the Wildfire–AQI Relationship?

The third research question examined whether the relationship between wildfire activity and air quality across California counties is moderated by social vulnerability. Social vulnerability was measured using the CDC's Social Vulnerability Index (SVI), which was mean-centered to aid interpretation of the interaction term. RPL_THEME1 of SVI was used; it measures a community's capacity to prepare for, respond to, and recover from hazardous events, with higher values indicating greater vulnerability. With the repeated observations nested within counties in the data, a linear mixed-effects model with county-level random intercepts was used to account for unobserved baseline differences in AQI across counties.

The main-effects model showed that wildfire activity has a statistically significant association with air quality. A tenfold increase in burned area was associated with an estimated 1.38-point increase in mean AQI ($p < 0.01$), suggesting a modest but consistent effect of wildfire activity on air quality degradation. Social vulnerability was also independently associated with AQI—a one-unit increase in SVI corresponded to roughly a 14-point increase in mean AQI ($p = 0.02$), indicating that more vulnerable counties tend to have worse baseline air quality regardless of wildfire activity in a given year.

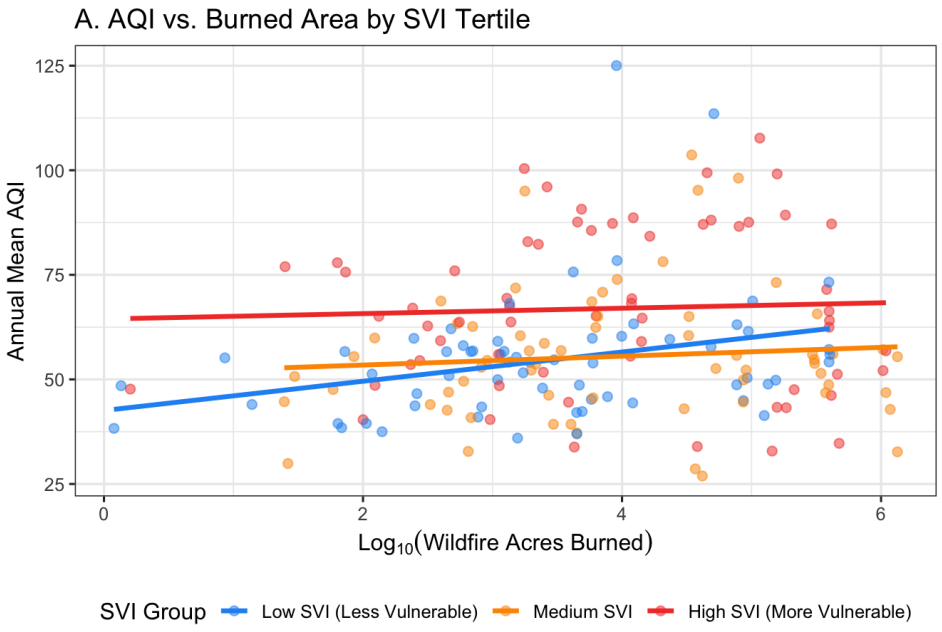


Figure 4: Model fit for AQI vs burned area by SVI tertile

Despite these main effects, the interaction between burned area and social vulnerability was not statistically significant ($p = 0.94$), and adding the interaction term did not meaningfully improve model fit over the simpler main-effects model ($\Delta AIC \approx 2$). Model-predicted values further supported this, showing nearly parallel trajectories across low, medium, and high SVI levels, which means the slope of the wildfire–AQI relationship does not meaningfully differ by social vulnerability. Overall, the hypothesis that more vulnerable counties would experience a stronger wildfire–AQI association was not supported by the data.

Discussion

Our findings support the hypothesis that larger wildfire-burned areas are associated with higher annual mean AQI at the county level. A 10-fold increase in burned area was associated with a 1.45-point increase in mean AQI, a shift that carries practical public health significance, as even moderate AQI increases can push counties across thresholds that trigger health advisories for sensitive populations. This result is consistent with Reid et al. (2016), who documented robust associations between wildfire smoke exposure and respiratory morbidity, and with Carreras-Sospedra et al. (2024), who estimated that the 2020 fire season alone caused thousands of premature deaths statewide.

From a practical standpoint, the effect of wildfire activity on AQI is relatively small for each increase in burned area, but it can add up during large fire seasons and lead to meaningful impacts on public health. In contrast, the effect of social vulnerability is much larger, suggesting that more vulnerable communities consistently experience worse air quality overall. While the estimated effect size of wildfire activity is statistically significant but modest, the magnitude of the SVI effect is substantially larger, indicating that baseline

inequality contributes more to overall air quality disparities than wildfire sensitivity. There are also some limitations: the sample size is not large enough to detect more subtle interaction effects, and using county-level SVI may hide important differences within counties.

The significance of social vulnerability as an independent predictor with more vulnerable counties showing AQI approximately 14.4 points higher regardless of fire activity suggests that air quality burden reflects not only wildfire exposure but also structural socioeconomic disadvantage. This aligns with Vargo et al. (2023), who found that minority, low-income, and limited-English communities experienced the largest increases in heavy smoke exposure between 2011 and 2021, and with Jung et al. (2024), who found that highly vulnerable counties show disproportionately greater increases in unhealthy air days.

Consistent with the model results, we find no evidence that social vulnerability modifies the wildfire–AQI relationship. Rather, the results suggest that the marginal effect of wildfire activity on AQI is relatively consistent across counties, regardless of vulnerability level. This pattern is consistent with prior literature indicating that wildfire smoke can disperse over large geographic areas, resulting in regionally distributed air quality impacts rather than localized effects (Jaffe et al., 2020). At the same time, the observed association between higher social vulnerability and elevated baseline AQI aligns with environmental justice research demonstrating that disadvantaged communities are disproportionately exposed to poorer environmental conditions (Reid et al., 2016; Vargo et al., 2023). Importantly, these findings suggest that social vulnerability shapes baseline exposure to air pollution rather than altering sensitivity to wildfire-driven changes in air quality.

In LA County, we found a significant relationship between burned acreage and days with an AQI over 100. This aligns with the findings of Jaffe et al. (2020), who reported an association between fire activity and PM2.5 levels in California. While encouraging, this model still had several limitations. Since there were only 48 data points, it was not as robust as we would have liked. The model was very overdispersed, introducing uncertainty in the errors and p-value.

If time allowed, additional analyses including daily data and the years 2017, 2019, and 2021 would have improved this aspect of our analysis considerably. Since this analysis did not use SVI data, those years have available data, but the amount of data cleaning and wrangling that was required was not in the scope of this project. Additionally, start and end dates for fires were provided, which could have given a more detailed look at exactly how fires were influencing the number of unhealthy AQI days.

Overall, we had significant limitations with our data. Temperature, drought severity, and percent urban land cover were unavailable in the merged dataset, meaning the burned area coefficient may capture some of their effects. Uneven AQI monitoring coverage across rural counties may also bias associations toward the null. SVI data is only published every two years, which limited the number of years we could include. Combining this with fire data starting in 2015, there were only four years that we had consistent data for. Daily data for

AQI and wildfire existed, while SVI and asthma data were limited to a yearly basis for each county. Additionally, the asthma data was not very reliable and had large deviations, so it was not used significantly in our analyses. These findings are most applicable to California counties with sufficient AQI monitoring during the study period and may not generalize to regions with different wildfire regimes or atmospheric conditions. Future work should incorporate additional environmental variables and higher-resolution temporal data to better capture wildfire impacts.

Conclusion

This study demonstrates that wildfire activity is a consistent driver of air quality degradation across California counties, while social vulnerability is associated with substantially higher baseline pollution levels. Contrary to expectations, we find no evidence that social vulnerability modifies the wildfire–AQI relationship, suggesting that wildfire impacts are broadly distributed rather than concentrated among more vulnerable populations. These findings highlight that environmental inequality is driven more by persistent baseline conditions than by differential sensitivity to wildfire exposure. Addressing air quality disparities therefore requires both wildfire mitigation and targeted efforts to reduce structural inequalities in vulnerable communities.

Attribution

Every group member contributed to conceptualization, methodology, software, validation, and analysis. Cecilia and Abrao provided resources for background information and discussion, and Flannery was the main contributor to data curation. Cecilia wrote the original draft of the introduction and background, and Abrao wrote the original draft of the methodology. Flannery created the initial data visualizations. All group members contributed to review and editing of the report and presentation equally.

Github repository link: https://github.com/fmcnair/ENV710_Group12

Bibliography

Carreras-Sospedra, M., Zhu, S., MacKinnon, M. et al. (2024). Air quality and health impacts of the 2020 wildfires in California. fire ecol 20, 6. https://doi.org/10.1186/s42408-023-00234-y

Cleland, S. E., Paul, N., Coker, E. S., & Henderson, S. B. (2025). The co-occurrence of wildfire smoke and extreme heat events in British Columbia, 2010–2022: evaluating spatiotemporal trends and inequities in exposure burden. ACS ES&T Air, 2(3), 319-330.

Do, V., Chen, C., Benmarhnia, T., & Casey, J. A. (2024). Spatial heterogeneity of the respiratory health impacts of wildfire smoke PM2. 5 in California. GeoHealth, 8(4), e2023GH000997

Hadiwijoyo, E., Rijal, H.B., Abdullah, N. (2025). Cascading Impacts of Wildfire Emissions on Air Quality, Human Health, and Climate Change Based on Literature Review. Fire 8, 471. https://doi.org/10.3390/fire8120471

Jaffe, D. A., O'Neill, S. M., Larkin, N. K., Holder, A. L., Peterson, D. L., Halofsky, J. E., & Rappold, A. G. (2020). Wildfire and prescribed burning impacts on air quality in the United States. Journal of the Air & Waste Management Association, 70(6), 583–615. https://doi.org/10.1080/10962247.2020.1749731

Jung, J., et al. (2024). Advancing the community health vulnerability index for wildland fire smoke exposure. Science of the Total Environment, 906, 167834. https://doi.org/10.1016/j.scitotenv.2023.167834

Rappold, Ana G., et al. "Community vulnerability to health impacts of wildland fire smoke exposure." Environmental Science & Technology 51.12 (2017):6674-6682 DOI: 10.1021/acs.est.6b06200

Reid, C. E., et al. (2016). Critical review of health impacts of wildfire smoke exposure. Environmental Health Perspectives, 124(9), 1334–1343. https://doi.org/10.1289/ehp.1409277

Tillakaratne, R., Hoshiko, S., Rosenberg, A., Hayashi, T., Buckman, J. R., & Rappold, A. G. (2023). Wildfires and the Changing Landscape of Air Pollution-Related Health Burden in California. American Journal of Respiratory and Critical Care Medicine, 207(7), 887-898. https://doi.org/10.1164/rccm.202207-1324OC

Vargo, J., et al. (2023). Social Vulnerability in US Communities Affected by Wildfire Smoke, 2011 to 2021. American Journal of Public Health, 113(7) https://doi.org/10.2105/AJPH.2023.307286

Zheng, Z., Murray, A. T., Sweeney, S. H., & Arabadjis, S. D. (2025). Social vulnerability gaps: Dynamics of the unsheltered homeless and wildfire hazards. Applied Geography, 180, 103648.

Data sources:

1. CDC/ATSDR (Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry). 2022. Social Vulnerability Index (SVI): Data and Documentation Download. U.S. Department of Health and Human Services, Atlanta, Georgia, USA. https://www.atsdr.cdc.gov/place-health/php/svi/svi-data-documentation-download.html. Accessed March 2026.

2. CAL FIRE (California Department of Forestry and Fire Protection). 2024. California Fire Perimeters (All). California Natural Resources Agency Open Data Portal, Sacramento, California, USA. https://catalog.data.gov/dataset/california-fire-perimeters-all-5b4ee. Accessed March 2026.

3. CDPH (California Department of Public Health). 2024. California Breathing: County Data. California Environmental Health Investigations Branch, Sacramento, California, USA. https://www.cdph.ca.gov/Programs/CCDPHP/DEODC/EHIB/CPE/Pages/CaliforniaBreathingData.aspx. Accessed March 2026.

4. U.S. EPA (U.S. Environmental Protection Agency). 2024. Air Quality System (AQS) Data Mart: Pre-generated Data Files. Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina, USA. https://aqs.epa.gov/aqsweb/airdata/download_files.html#AQI. Accessed March 2026.