Technical Appendix

All code and data used in the model are publicly available from xxx. This document is intended to complement the codebase by explaining assumptions, discussing modeling decisions, and providing information on data sources.

# Generating microdata

A microdata set from the 2021 ACS was obtained through IPUMS USA with select variables of interest, including county and household identifiers, and sociodemographic variables. Details on how ACS data is collected can be found here. We added a rural/urban flag using county designations defined [here](https://www.counties.org/sites/main/files/file-attachments/2020-june3-countycaucusesinfographic-4-final.pdf). Suburban counties were categorized as rural.

Current asthma prevalence data by age group and county was obtained from the [California Department of Public Health](https://www.cdph.ca.gov/Programs/CCDPHP/DEODC/EHIB/CPE/Pages/CaliforniaBreathingCountyAsthmaProfiles.aspx), sourced from the 2019-2020 California Health Interview Survey. Missing data was imputed using the age-specific California average for age group 0-4 (91% missing), and age-specific population-weighted residual average for age group 5-17 (53% missing). Population weights were derived from the 2010 census, which was the most recent data available.

We recognize that asthma is not evenly distributed within counties. A major factor is proximity to sources of pollution, which are typically in or near lower-income neighborhoods. We therefore used data on poverty (household income as a percentage of the federal poverty line) to adjust the probability of having asthma for each individual in the dataset (Table ##). A binary asthma variable was created using these probabilities.

Table ##. Poverty and asthma prevalence

|  |  |  |
| --- | --- | --- |
| **Household income** | **Proportion of sample** | **Risk ratio for asthma** |
| 200% or more of the poverty line | 0.75 | (reference) |
| 100% or more but less than 200% of the poverty line | 0.14 | 1.32 |
| Less than 100% of the poverty line | 0.11 | 1.62 |

Lastly, for individuals who were assigned as having asthma, we defined an asthma control variable, by sampling from the vector of asthma control states given the probabilities of having each level of control. These control probabilities were calculated during model calibration and are described elsewhere.

# Data sources

## Census data

The data from which the model population is created was from the 2021 American Community Survey, obtained from [IPUMS USA](https://usa.ipums.org/usa/about.shtml), who provide and maintain a database of public use microdata drawn from numerous federal censuses and national surveys. IPUMS does not identify places with a population fewer than 100,000. As a result, the California data we extracted include individuals from 35 of California’s 58 counties. These 35 counties are representative of 96.4% of the California population. A list of extracted variables is available in the data dictionary available on the project repository.

## Asthma prevalence

Data on current asthma prevalence was obtained from the [CDPH California Asthma Dashboard](https://www.cdph.ca.gov/Programs/CCDPHP/DEODC/EHIB/CPE/Pages/CaliforniaBreathingCountyAsthmaProfiles.aspx), which reported California Health Interview Survey data, on 6/21/2023. Detailed notes on the data can be found [here](https://www.cdph.ca.gov/Programs/CCDPHP/DEODC/EHIB/CPE/CDPH%20Document%20Library/Notes_About_the_Data_ADA.pdf). The dataset included asthma prevalence by age group and county.

Across all age groups, county-specific prevalence ranged from 0.3% to 33.4%, with an overall California prevalence of 8.7%. Prevalence was mostly higher in suburban and rural counties and was especially high among 5–17-year-olds in Butte County (33.4%), and those over 65 in Solano (23.8%) and Kings (21.4%) counties. (add one of the figures)

As with the census data, prevalence data were suppressed when sample sizes were small in order to protect confidentiality. Younger age groups and smaller counties were therefore missing. The youngest age group of 0-4-year-olds had data for only five counties. We therefore used the California prevalence for that age group (3.3%) when the county-specific prevalence was not available. The next age group of 5-17-year-olds also had 53% missing data. Given greater data availability, we imputed the “residual” average for the missing counties weighted by the population size of 5-17-year-olds in each county. Population size by age were obtained from the 2010 Census, via Social Explorer tables (SE\_T008), to compute county population weights.

## Fire and smoke

We chose the 2018 Camp Fire as given wide availability of evidence for both the smoke pollution it caused as well as health outcomes the event led to. Daily average PM2.5 was selected as the smoke variable. While PM2.5 is among the most dangerous pollutants associated with wildfires, there are limitations to its use as a predictor variable. PM2.5 is not a single substance, rather it refers to a mixture of different chemical compounds. The composition of wildfire PM2.5 greatly differs from ambient pollution and is considered to contain more toxic substances. Measurements of PM2.5 concentration in the air does not provide information about its composition, and therefore does not allow to distinguish differences in potential health impacts. However, the Camp Fire caused substantial increases in PM2.5 concentration in the areas it affected, therefore any PM2.5-related health outcomes over the period of the fire could reasonably be attributed to wildfire smoke rather than ambient pollution.

Daily average PM2.5 measurements between 2018-2019 were obtained from the [Environmental Protection Agency’s Air Data portal](https://www.epa.gov/outdoor-air-quality-data/download-daily-data). Improbable measurements of non-positive PM2.5 concentration were removed as potential measurement errors (0.6% of all observations). For counties with multiple monitors, measurements from multiple monitors were averaged. We then calculated a background PM2.5 value for each county by calculating the median daily PM2.5 during 09/01/2018 and 11/01/2018 (this date range was selected after visualizing trends in PM2.5 over time). To calculate the specific impact of the Camp Fire on PM2.5 pollution, we subtracted this background pollution value from the daily PM2.5 for fire days. Dates outside of the fire period were assigned 0 for fire-specific PM2.5.

# Model Calibration

1. **Weekly transition probabilities** were calibrated to empirical values on prevalence of each asthma control state. (paste the blurb here)
2. **Risk ratios for smoke exposure on transition** probabilities matched empirical values on asthma healthcare use with/without wildfire smoke. We utilized the risk ratios estimated by [Wu et al.](https://bmcpediatr.biomedcentral.com/articles/10.1186/s12887-019-1530-7) for the association between smoke exposure and asthma exacerbation. While their dependent variable (exacerbation) is not necessarily a direct representation of asthma control, it was the closest available estimate we could find, and the study design matched well with our model in terms of measurement timepoints. We tested whether these risk ratios generated asthma healthcare use outcomes comparable to empirical data by calculating the modelled increases in healthcare use (urgent/outpatient care, emergency department visits, and inpatient hospitalizations) during cycles in which smoke was present. The calculated increases were similar to those reported by [Hutchinson et al.](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6038982/) and [Malig et al.](https://www.sciencedirect.com/science/article/pii/S004896972102578X?via%3Dihub), and no further calibration was done.

Clinical states are heavily associated with exacerbations and healthcare use. Clear on the timeframes, and what the findings of the studies are.

Add the Chinese study to the table, indicate timeframe of exposure, exact outcome, study settings. Show that they’re consistent.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Type of encounter** | **None** | **OCS burst** | **Urgent care / Outpatient visit** | **Emergency department visit** | **Hospitalization** |
| **Simulated proportion without smoke** | 0.796 | 0.090 | 0.078 | 0.017 | 0.019 |
| **Simulated proportion following 1 week of smoke days (cutoff 35)** | 0.753 | 0.106 | 0.081 | 0.033 | 0.027 |
| **Simulated % change** | *-5.4%* | 18% | 3.8% | 94% | 42% |
| **Empirical evidence from Malig et al.a** | n/a | n/a | n/a | 112% | 53% |
| **Empirical evidence from Hutchinson et al.b** | n/a | n/a | ~30% | 112%\* | ~50% |

aStudy of 2017 North San Francisco Bay wildfires. Reports rates of ED visits and hospitalizations during the fire period (10/9/17 – 10/18/17) and non-fire period (10/1/15 – 10/18/15; 10/1/16 – 10/18/16; 10/1/17 – 10/8/17).

bStudy of 2007 San Diego wildfires. Reports risk ratios of ED visits, hospitalizations, and outpatient visits during periods of fire exposure (10/22/2007 – 10/26/2007) and compared with occurrences during six 5-day periods matched on the day of the week starting 3-9 weeks before the exposed period.

\*The two studies calculated identical increases in emergency department visits.

**3) Risk ratios for age on transition probabilities** were calculated from [Bloom et al](https://thorax.bmj.com/content/thoraxjnl/73/4/313.full.pdf).’s 2017 study of asthma in the UK population. The study evaluated rates of asthma exacerbation across age groups using electronic health record data. As above, this is closely related with asthma control, thus we directly applied them in our model.

|  |  |  |  |
| --- | --- | --- | --- |
| **Age group** | **Reported rate per 10 person-years (95% CI)** | **Calculated probability (weekly)** | **Calculated risk ratio** |
| **Under 5** | 4.27 (4.18 – 4.38) | 0.00818 | (ref) |
| **5 to 17** | 1.48 (1.47 – 1.50) | 0.00284 | 0.348 |
| **18 to 54** | 3.22 (3.21 – 3.24) | 0.00617 | 0.755 |
| **55 and above** | 9.49 (9.37 – 9.42) | 0.01808 | 2.211 |