**Estimating wildfire smoke PM2.5 exposure:**

Wildfire smoke PM2.5 data come from Childs et al. 2022, which produced 10km2 gridded estimates for the entire US for all days, 2006-2020.1 In brief, their methods involved determining which days had wildfire smoke plumes overhead based on satellite imagery. For such “smoke days,” PM2.5 anomalies from EPA ground stations were attributed to wildfire smoke PM2.5, and the smoke-attributable PM2.5 value was calculated as the difference between the anomalous value and the 3-year non-smoke day median. A machine learning model was used to predict ground station values, including data on meteorological measures, aerosol optical depth predictions, HYSPLIT trajectory points, and topological data, among other sources. After training, this ML model was used to generate gridded, smoke-attributable PM2.5 predictions for the entire country. These data have already been used to measure exposure in environmental epidemiology studies such as Wen & Burke 2022.2

Population-weighted aggregation was used to take the daily, 10km2 gridded smoke PM2.5 estimates from Childs et al. to the county-month level. Population data came from WorldPop’s 2013 estimates, which provide values of population density at the 10km2 level. The proportion of each county that overlapped with a grid cell was calculated, population densities were combined with these areas to determine the total population of each county-grid cell overlap, and then the mean smoke PM2.5 estimate across all overlapping grid cells was calculated, weighted by grid cell population. Population weighting was used because counties often cover large tracts of land and do not have an evenly distributed population, so area-weighting estimates may misrepresent the population’s true exposure to smoke PM2.5 or other environmental variables.

**Mortality and population data:**

Mortality data were obtained from the CDC’s Wide-ranging Online Data for Epidemiologic Research (WONDER) site’s 1999-2000 multiple cause of death data.3 WONDER mortality data is based on death certificates of US residents, with information on primary cause of death, additional non-primary causes, and demographic data. Aggregated versions of these data are publicly available without a DUA, but a DUA is necessary to access the granular counts, including cells with fewer than 10 events.

The primary outcome in this study is all-cause mortality, as reported in the CDC WONDER database, which captures all death certificates from US residents. All-cause mortality was assessed using ICD10 codes A00-Z99. Cause-specific mortality was determined using ICD10 codes I00-I99 for cardiovascular mortality; H62-H67 and J00-J99 were used for respiratory mortality.

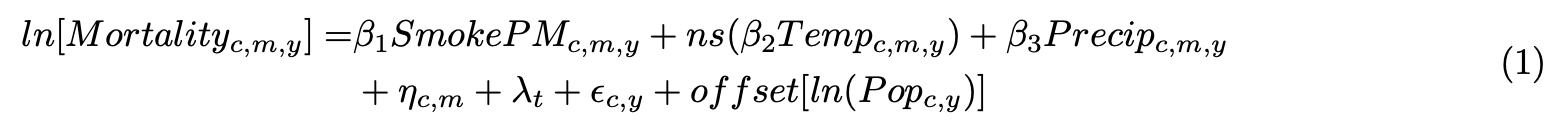
Data from WONDER were available with county-level population estimates stratified by age, sex, and race/ethnicity, but lacked data on marital status. The American Community Survey 5-year data (ACS-5) begins in 2009 and includes data on marital status, and therefore was used on conjunction with the WONDER data. The “B12002” table of the ACS-5 was used to gather data on marital status across age categories. Mortality and marital status data were aggregated at the county-month level with stratifications by age, sex, race/ethnicity, and marital status for the years 2009-2020, excluding the months after February in 2020 to avoid disruptions due to the COVID-19 pandemic. This roughly decade-long period included many major wildfire events and captured the trend of increasing wildfire smoke exposure in the US, which is expected to continue in the coming decade.

**Environmental control data:**

Monthly mean temperature and precipitation will be used as important covariates that would not otherwise be accounted for in the TWFE model. Monthly mean temperature and precipitation data at the 4km2 level will be obtained from Oregon State’s PRISM Climate group and aggregated to the county level by population-weighting.. Other potential confounders do not need to be controlled for due to the nature of the analytic approach.

**Estimating the effect of wildfire smoke PM2.5 on mortality:**

Our models utilized a quasi-Poisson regression with two-way fixed effects. Quasi-Poisson regression was chosen because our total mortality is a count variable, though the model results can be interpreted as an effect on the mortality rate due to the offset term. We did not choose negative binomial regression because, although both models are capable of handling overdispersed data, negative binomial regression gives outsized weight to smaller counties, while quasi-Poisson regression weights counties by size more evenly.4 The regression is shown in equation 1:



This regression models the natural logarithm of mortality counts for each county *c*, month *m*, and year *y*. The county-month fixed effect ˙˙ƒ

*ηc,m* creates dummy variables for each county-month in our sample (in other words, 3,083 counties \* 12 months = 36,996 intercepts). We chose this fixed effect instead of the more common district-only fixed effect because different counties may have different seasonal effects for variables unmeasured in our regression. This fixed effect means the model compares each county-month’s mortality counts to themselves from year to year, theoretically eliminating the need to control for both measured and unmeasured year-invariant county-month-level confounders such as socioeconomic status, urbanicity, and so forth.

The year fixed effect *ƛt* creates dummies for each year of our sample (11 dummies total for years 2009-2020), controlling for nationwide secular trends. *SmokePMc, m, y* represents the mean daily wildfire smoke-attributable PM2.5 pollution in county *c* during month *m* of year *y*, and is our main variable of interest. We used no lag terms, so this regression models the effect of same-month smoke PM2.5 on mortality. *Tempc,m,y*represents the mean monthly temperature of each county, and a natural cubic spline with 3 degrees of freedom is applied to account for nonlinear effects of temperature on mortality. Similarly, *Precipc,m,y* represents the mean precipitation for each county, modeled linearly. The offset term *ln(Popc,y)* represents the total population of each county *c*  in each year *y*, and it is included so that the model results can be interpreted as an effect on mortality rates instead of on raw mortality counts. εc,m,y represents the error term. β*1* represents the average effect of an additional µg/m3 of average monthly smoke PM2.5, though it must be exponentiated to be interpretable as a risk ratio. The model was weighted by county population size. Robust standard errors were used, and they were clustered at the county level to account for within-unit correlation. Modeling was done using the fixest package in R 4.2.0. Tests for significance were two-tailed.

This model’s county-month and year fixed effects cannot control for confounders that vary across both time and county. We chose to directly adjust for temperature and precipitation because they clearly vary across both county and time, and are known to be associated with wildfire risk, ambient PM2.5, and mortality.5–8 Other potential confounders are not directly measured, and represent a possible source of bias.

While the offset term does allow the model coefficients to be interpreted as affecting the mortality rate instead of the raw death counts, and this effect on the mortality rate is not age-adjusted, strictly speaking. However, the county-month fixed effect accounts for county-level population age structures, and the year fixed effect accounts for nationwide trends in aging. These combined effects minimize the concern of confounding by population age structure, because in order for it to be a confounder, county age structure would have to vary year-to-year in a way that is correlated with changes in smoke PM2.5 and is not captured in national trends, which is not highly plausible.

1. Childs ML, Li J, Wen J, et al. Daily Local-Level Estimates of Ambient Wildfire Smoke PM 2.5 for the Contiguous US. *Environ Sci Technol*. 2022;56(19):13607-13621. doi:10.1021/acs.est.2c02934

2. Wen J, Burke M. Lower test scores from wildfire smoke exposure. *Nat Sustain*. 2022;5(11):947-955. doi:10.1038/s41893-022-00956-y

3. Multiple Cause of Death 1999-2020. Accessed February 28, 2023. https://wonder.cdc.gov/wonder/help/mcd.html

4. Ver Hoef JM, Boveng PL. QUASI-POISSON VS. NEGATIVE BINOMIAL REGRESSION: HOW SHOULD WE MODEL OVERDISPERSED COUNT DATA? *Ecology*. 2007;88(11):2766-2772. doi:10.1890/07-0043.1

5. An H, Gan J, Cho S. Assessing Climate Change Impacts on Wildfire Risk in the United States. *Forests*. 2015;6(12):3197-3211. doi:10.3390/f6093197

6. Basu R. High ambient temperature and mortality: a review of epidemiologic studies from 2001 to 2008. *Environ Health*. 2009;8(1):40. doi:10.1186/1476-069X-8-40

7. Basu R. Relation between Elevated Ambient Temperature and Mortality: A Review of the Epidemiologic Evidence. *Epidemiol Rev*. 2002;24(2):190-202. doi:10.1093/epirev/mxf007

8. Tai APK, Mickley LJ, Jacob DJ. Correlations between fine particulate matter (PM2.5) and meteorological variables in the United States: Implications for the sensitivity of PM2.5 to climate change. *Atmos Environ*. 2010;44(32):3976-3984. doi:10.1016/j.atmosenv.2010.06.060