# Ashes to Ashes: The Lifelong Consequences of Early-Life Wildfire Exposure*

Samuel Arenberg† Seth Neller‡

October 13, 2022

This paper assesses the impact of early-childhood exposure to wildfire smoke on longevity, economic achievement, human capital, and disability in mid-to-late adulthood. To do so, we leverage novel data from mid-20th century California fire atlases and smoke dispersion modeling, combined with restricted-use Census Bureau records and administrative mortality data. We find that moving from the 25$^{th}$ to 75$^{th}$ percentile of early-life wildfire smoke exposure results in 46 life years lost per 1,000 persons due to shortened adult lifespan and results in unfavorable changes to several indices of adult well-being. Our findings suggest that warming temperatures, which exacerbate the duration and intensity of wildfire seasons, are already affecting the life cycles of exposed children through increased smoke exposure.

*First version: October 30, 2021. This version: October 13, 2022. We are grateful to Marika Cabral, Mike Geruso, and Dean Spears for their guidance and support. We also thank Martha Bailey, Sandy Black, Bokyung Kim, Niklas Kroner, Yonah Meiselman, Rich Murphy, Sheila Olmstead, Jinyeong Son, Cody Tuttle, Anjali Verma, Tom Vogl, and Sangita Vyas, as well as seminar attendees at various seminars and conferences for helpful comments and suggestions. We are exceptionally grateful to Martha Bailey for providing the crosswalk from the Numident string names into county codes and to Karin Johnson for administering this project. Any views expressed are those of the authors and not those of the U.S. Census Bureau. The Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers have reviewed this information product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 2272. (CBDRB-FY21-P2272-R9133, CBDRB-FY21-P2272-R9219).

†National Bureau of Economic Research. Email: samuel.arenberg@gmail.com

‡Department of Economics and Boyd Center for Business and Economic Research, University of Tennessee. Email: sneller@utk.edu

# 1 Introduction

It is now well established that air pollution increases contemporaneous mortality of both the very old and very young, two age groups most sensitive to their environments. However, substantially less is known about a related question, which we address in this paper: how do the harms of exposure to air pollution at very young ages manifest in reduced well-being and longevity in later adulthood? Given that long-run effects frequently comprise a large fraction of the damages from early-life shocks (Currie and Almond, 2011; Almond et al., 2018), understanding the ramifications of childhood pollution exposure across the life cycle is vital in assessing its total costs. This paper provides the first evidence that increases in ambient air pollution in early life increase mortality at older ages and is also the first——to our knowledge——to quantify the effects of childhood pollution shocks on measures of earnings and disability past early adulthood (through age 59).

To determine the long-run effects of early-life exposure to air pollution, we study wildfires, the only major emissions source of fine particulate matter ($PM_{2.5}$) in the United States that is growing over time and one that has generated approximately 29% of all fine particulate pollution in recent years.[1] Additionally, it is expected that wildfire pollution will *continue* to grow: the United States Forest Service predicts that 1 °C increase in average temperature will result in a 200-600% increase in acres burned in the Western United States (Vose et al., 2012). Given that experts predict global temperatures will increase by at least 2 °C over the global baseline by 2050 (IPCC, 2021), the prospect of increased pollution from wildfires is all but certain.

There are two key challenges to conducting a long-term analysis of childhood air pollution. The first is that pollution has been reliably measured only since the mid-to-late 1960s, with the first meaningful regulation occurring with the Clean Air Act Amendments of 1970. This lack of historical data and sources of exogenous variation has limited researchers' ability to evaluate the ramifications of early-life pollution exposure over the life cycle. We overcome this challenge by examining wildfires occurring in California, a source of pollution which has been tracked reliably since at least 1930. Additionally, wildfire smoke exposure, which is a function of fire timing, acres burned, wind direction, and wind speed, provides potentially exogenous variation for an analysis of long-term outcomes. In order to supplement these wildfire records and generate measures of smoke exposure, we first hand-gather dates in which the wildfires occurred using hundreds of historical newspaper articles. We then utilize wildfire pollution modeling tools to generate historical measures of wildfire smoke exposure. These modeled smoke measures are used by the United States Forest Service to predict wildfire smoke and compare well to patterns generated by satellites (Larkin et al., 2009). In this paper, we perform additional exercises to further validate this modeled smoke measure and find that it strongly predicts pollution exposure in a modern context.

The second challenge in assessing the long-run impacts of early-life shocks is linking childhood circumstances to later-life outcomes. To address this difficulty, we obtain restricted-use administrative data from the Social Security Administration and U.S. Census Bureau. In addition to providing

[1] Source: Author calculations using National Emissions Inventory data from the Environmental Protection Agency. See Appendix Figure A1 for more detail.

comprehensive high-quality microdata, these data sets have two features that are ideally suited for our analysis. First, the administrative data provides exact city and state of birth, which allows us to accurately assign measures of wildfire smoke exposure to individuals. Second, these data also include unique identifiers which we use to link individuals—and their associated early-life exposure measures—to comprehensive death records (as applicable) as well as their responses in the 2000 and 2010 Decennial Censuses and 2005-2019 American Community Surveys.

We evaluate the long-term effects of exposure using a cohort distributed leads and lags model and find that the effects of childhood pollution exposure—which we define as exposure before age 5—last throughout the entire life cycle. Specifically, we find that moving from the 25$^{th}$ to the 75$^{th}$ percentile of exposure—roughly 5 additional months with *any* wildfire smoke—results in 1.7 additional deaths prior to age 55 per 1,000 persons, conditional on surviving past childhood.[2] Increases in mortality appear as early as age 30 and dissipate by age 80—summing the total effects between these two ages translates to 46 life years lost per 1,000 persons. These reductions in adult longevity exceed the number of life years lost due to infant mortality attributable to wildfire pollution. This further underscores the fact that quantifying long-run effects is an essential exercise in assessing the costs of childhood harms.

To obtain insight into how air pollution affects well-being in mid-adulthood and provide evidence regarding potential channels for the estimated reductions in longevity, we also consider the effects of heightened wildfire smoke during childhood on non-mortality outcomes at ages 35 through 59. Because of the wide range of outcomes available to us in our data—and because early-life shocks have been shown to affect nearly every facet of adult life—we standardize outcomes (à la Kling et al., 2007) and combine them into three indices covering economic achievement, educational attainment, and disability. We also specifically consider "headline" outcomes within each group that typify our results and are more easily interpreted. We find that moving from the 25$^{th}$ to 75$^{th}$ percentile of smoke exposure decreases our index of economic achievement by 1.45% of a standard deviation (%-SD), headlined by an $890 decrease in annual earned income (defined as wages and self-employment income). Consistent with our economic results, we find that the same level of exposure decreases our educational attainment index by 1.35%-SD, including a 0.40 percentage point decrease in high school completion. Finally, we find an adverse change of 0.80%-SD in our index of disabilities. Our most notable finding within this group of outcomes is that moving from the 25$^{th}$ to 75$^{th}$ percentile of smoke exposure increases the likelihood of difficulty remembering, concentrating, or making decisions by 0.30 percentage points, supporting the idea that early-life pollution exposure impacts cognition and mental health throughout the entire life cycle. Taken together, we conservatively estimate that the average child in our sample sustains approximately $22,000 of discounted damages in lost earnings and lost life years due to wildfire smoke.

To demonstrate that our results are driven by smoke exposure—rather than proximity to the

[2] Results are conditional on surviving past childhood because it is necessary for an individual to obtain a Social Security number to be included in our data, a process that did not typically occur until adolescence for the cohorts born in our sample. Also of note: we feature survival until age 55 as a main outcome as that is the latest age that we observe all cohorts in our sample.

actual wildfires themselves—we control for nearby fires during childhood. Inclusion of these controls has almost no effect on our estimates, suggesting that the impacts that we find relate to pollution damage, rather than long-run effects of trauma or negative economic shocks due to natural disaster exposure. Additionally, we show that our estimates are stable across a range of robustness tests, including adding controls for climate and other time-varying factors as well as different functional forms and specifications. We also consider the role of mortality selection by including bounded estimates in a similar spirit as Lee (2009) and find that wildfire smoke's differential effects on longevity do not materially affect estimates of other long-run outcomes. Additionally, we consider alternate estimation methods to evaluate the impact of weighting issues endemic to models that utilize two-way fixed effects (e.g., Callaway and Sant'Anna, 2021; de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021a; Sun and Abraham, 2021) and find that results are qualitative similar when estimated using other methods.

To better understand how childhood smoke exposure affects long-run outcomes, we conduct a variety of supplemental analyses. First, to further understand how the dynamics of childhood wildfire smoke exposure evolve as individuals grow older, we consider heterogeneity by age at survey. We find economically larger effects—both in levels and in relation to the mean—on earnings and cognitive difficulty at older ages. This finding suggests that evaluations of the impacts of early-life air pollution—and potentially other childhood shocks—should consider later-life dynamics to capture the total costs. Next, to better understand how wildfire pollution exposure affects the distribution of outcomes, we consider heterogeneity across the economic index and wage distributions. We find that early-life exposure to wildfire pollution causes in a leftward shift of the entire distribution, suggesting that both extensive-margin labor supply and earnings (conditional on employment) are affected. Finally, we investigate occupational and neighborhood characteristics to provide insight into the ways in which economic circumstances may contribute to reduced longevity. We find that individuals affected by wildfire smoke during childhood select occupations and neighborhoods that have higher mortality risk, providing suggestive evidence that economic factors could be driving mortality results, particularly those in early adulthood.[3]

Our findings that early-life exposure to wildfire smoke pollution reduces longevity and adversely affects economic achievement, educational attainment, and disability in mid-to-late adulthood contribute to three distinct areas of economic research. First, this paper contributes to our understanding of how circumstances in early childhood affect adult longevity. To date, nearly all of our understanding in this area comes from studies showing that expanded access to the social safety net while young decreases mortality at later ages (e.g., Aizer et al., 2016; Bailey et al., 2020; Goodman-Bacon, 2021b). To the best of our knowledge, this is the first causal research to assess the longevity impacts of childhood exposure to ambient air pollution from any source.[4] These

[3] Additionally, we perform further analyses to understand how local economic shocks from smoke exposure (Borgschulte et al., 2020) could affect long-run outcomes, explore effect heterogeneity by person and place characteristics, and assess the impact of childhood air pollution exposure on other long-run outcomes that do not fall into the categories of economic achievement, educational attainment, or disability.

[4] Notably, we are also unaware of any research displaying non-causal *associations* between childhood air pollution and adult mortality. This is likely due to the substantial difficulties in locating historical pollution measures and longitudi-

longevity effects are important to measure, as they result in a greater number of life years lost than those lost from infant deaths and also comprise over half of the total damages from smoke exposure measured in this paper.

Second, we expand the literature examining the long-run effects of early-life air pollution on adult outcomes. This work has largely focused on the impact of the long-run effects of the Clean Air Act Amendments ("CAAA") on earnings around age 30 (Isen et al., 2017b; Colmer et al., 2022) and on educational attainment (Colmer and Voorheis, 2020; Voorheis, 2017).[5] This paper increases our understanding of the long-run effects of childhood pollution exposure in two primary ways. First, we causally assess the impacts of early-life air pollution substantially past age 30—in our setting, we examine impacts on ages 35 through 59—and are the first paper to do so.[6] This is a meaningful contribution because early-life impacts in some settings can be detected at younger ages and then "fade" over time (see Currie and Almond, 2011; Almond et al., 2018, for examples), while it is also possible for childhood harms to *increase* proportionally as individuals age. Understanding that childhood air pollution shocks grow proportionally worse as individuals age—as we find in this paper—is important when assessing the cumulative costs of early-life harm from air pollution and may also have implications for other childhood health shocks. The second way this paper complements existing research is by examining a different—and highly relevant—source of pollution. While the pollution from industrial and vehicular sources studied in this literature have been falling for many years, pollution from wildfire smoke has been increasing—and likely will continue to increase—over time.[7] Given the growing prevalence of wildfires and evidence from medical research suggesting that particulates from wildfires are potentially more harmful to human health than pollution from other sources, understanding the harm done by wildfire pollution is increasingly important.[8]

Third and finally, this study contributes to our rapidly growing understanding of the effects of pollution from fires on human health. Thus far, this literature has concentrated primarily on the contemporaneous effects of fire-related pollution on mortality (Jayachandran, 2009; Miller et al., 2017), infant health (Rangel and Vogl, 2019), and labor market outcomes (Borgschulte et al., 2020). As emphasized above, this paper adds to the existing knowledge of these health effects by quantifying the long-run harms of wildfire smoke, which we find to be a substantial component of the

nally linking them to adult mortality outcomes, as discussed in previous paragraphs. Tangentially related are studies that combine cross-sectional variation in air pollution exposure with relatively short-run mortality measures to estimate the effects of air pollution on life expectancy (see Chen et al., 2013, for a notable example). However, because these papers use mortality measures that typically involve 1-to-5-year follow-up periods, they are still identifying the contemporaneous effects of air pollution, which is fundamentally different than the focus of this paper.

[5] An early contributor to this literature is Sanders (2012), who studies the long-run effects of recession-induced decreases in pollution on test scores. Additionally, there is also a small literature examining the impacts of the lead air pollution on educational attainment (Grönqvist et al., 2020) and crime/anti-social behavior (Aizer and Currie, 2019; Reyes, 2007, 2015).

[6] Within the medical literature, Russ et al. (2021) combine modeled pollution measures with longitudinal survey data ($n = 572$) and find an association between childhood air pollution and late-life cognitive decline.

[7] As demonstrated in Appendix Figure A1, wildfire smoke comprised over 29% of fine particulate matter pollution during the 2017-2020 period, more than the emissions from fuel combustion/generation, industrial activities, and vehicular use *combined*.

[8] See, for example, Aguilera et al. (2021a) and Aguilera et al. (2021b)

costs of these fires. Furthermore, insofar as pollution from these fires is expected to grow due to global increases in temperature (Moritz et al., 2012; Vose et al., 2012), we view this paper as an important contribution to our understanding of the costs of climate change (e.g. Jacob et al., 2007; Schlenker and Roberts, 2009; Graff Zivin and Neidell, 2014; Barreca et al., 2016; Hsiang et al., 2017).

# 2 Background and Conceptual Framework

## 2.1 Ambient Pollution and Childhood Development

Because this paper estimates the impact of air quality in early life on outcomes in late life, in this section, we first explain the biological influence of ambient pollution on human health. Further, we document several reasons why wildfire smoke may be more harmful than other forms of ambient pollution. Throughout, we focus on the role of fine particulate matter, or $PM_{2.5}$, because it is widely considered to be the most harmful to human health (Peeples, 2020).

$PM_{2.5}$ is particulate matter, either solid or liquid, that is less than 2.5 microns in diameter (for reference, a human hair is about 70 microns in diameter). It enters the body through inhalation into the lungs. Although some of the matter will be removed by normal clearance mechanisms, some of the matter will reach the bloodstream, where the particulates can cause a multitude of cardiovascular and respiratory problems. Although the precise pathways are not yet fully understood, a leading hypothesis is that many particulates carry or produce reactive oxygen species ("ROS"). Excesses of ROS can interfere with normal cellular processes and cause damage to tissues and organs. Such oxidative stress is associated with onset of diseases ranging from diabetes to atherosclerosis. $PM_{2.5}$ can even enter the gastrointestinal tract, causing imbalances in the intestinal microecology, which is known to affect a host of other systems, including the central nervous system. Further, fine particulate matter has been shown to weaken the immune system and cause increased susceptibility to infections months after initial exposure (Landguth et al., 2020).

The harm from fine particulate matter may be exacerbated in early life: young children breathe in more air relative to their body weight, they tend to spend more time outdoors, and their bodies are growing rapidly. Holm et al. (2021) summarizes the potential connection between wildfire exposure at young ages and outcomes at older ages, "[E]arly-life exposure may act by 'programming' fundamental metabolic, structural, and cell signaling mechanisms that may result in lifelong impacts." Furthermore, early-life and late-life health statuses are strongly correlated, suggesting that harms persist over the life cycle (Case and Paxson, 2010).

Finally, there is an emerging literature that suggests $PM_{2.5}$ from wildfire smoke may be worse for human health than other sources of ambient pollution.[9] One reason is that wildfire smoke has a greater concentration of ultra-fine particulates, which are less than 0.1 microns in diameter (Holm et al., 2021). These particles can penetrate deeper into the body than larger particulates, cross the blood-brain barrier, damage neural tissue, and impair brain development (Schraufnagel, 2020). Another reason is that wildfire smoke contains more polycyclic aromatic hydrocarbons, which

[9] Consult Holm et al. (2021) for a detailed overview of the potential health impacts of wildfire smoke on children.

induce greater oxidative stress. Wildfire smoke also contains more volatile organic compounds, many of which are understood to be carcinogenic (e.g., formaldehyde and benzene). Additionally, wildfire smoke has shown to contain meaningful amounts of heavy metals, such as lead, that have been linked to adverse long-term outcomes (California Air Resources Board, 2021; Grönqvist et al., 2020). The likelihood that wildfire smoke is more harmful than other forms of ambient pollution has been demonstrated by animal studies (Wegesser et al., 2010) as well as observational analyses of child hospitalizations (Aguilera et al., 2021a; Aguilera et al., 2021b). Further, a recent paper that examines the effects of wildfire smoke pollution on the labor market (Borgschulte et al., 2020) finds effects on extensive-margin labor supply that are 2-7 times larger than a comparison group of papers that utilize variation in other pollution sources. In summary, these studies support the idea that the composition of wildfire smoke particles may be particularly harmful for human health.

### 2.1.1 Conceptual Framework

To formalize the hypothesized impact of early-life pollution exposure on long-run outcomes and our reduced-form findings discussed in subsequent sections, we present a simple conceptual model.[10] This framework, which is visually represented by Appendix Figure A2, models the life cycle in four distinct periods. In the first period, childhood health ($h_1$) is determined as a function of genetic endowments ($h_0$) and inputs during early life ($I_1$), so that $h_1 = h_1(h_0, I_1)$.[11] Given its documented adverse health effects—discussed in detail above—air pollution in childhood is modeled as negatively affecting early-life inputs and health, such that $\partial h_1/\partial I_1 < 0$.

In the second period, individuals choose an amount of education ($e$), the quality of which is a function of childhood health: $e = e(h_1) = e(h_0, I_1)$. Because health is seen as a positive input into the educational process, childhood smoke exposure is predicted by the model to reduce educational attainment:

$$\frac{\partial e}{\partial I_1} = \frac{\partial e}{\partial h_1} \cdot \frac{\partial h_1}{\partial I_1} < 0.$$

After completion of their education careers, individuals progress into the labor market, the third period of the model. Because total earnings depend on both education and innate ability (physical and mental, both represented by health stock in our model), we model earnings ($\ell$) as:

$$\begin{aligned} \ell &= \ell(e(h_1), h_1) \\ &= \ell(e(h_0, I_1), h_0, I_1). \end{aligned}$$

With this structure, the impact of early-life pollution exposure is given by:

[10] What follows is similar in spirit to the frameworks presented by Isen et al. (2017b) and Colmer and Voorheis (2020).

[11] For purposes of this model, health is broadly defined to include physical health and measures of mental well-being, including cognitive ability and non-cognitive skills.

$$\frac{\partial \ell}{\partial I_1} = \underbrace{\frac{\partial \ell}{\partial h_1} \cdot \frac{\partial h_1}{\partial I_1}}_{\substack{\text{Lower earnings due} \\ \text{to poorer health}}} + \underbrace{\frac{\partial \ell}{\partial e} \cdot \frac{\partial e}{\partial h_1} \cdot \frac{\partial h_1}{\partial I_1}}_{\substack{\text{Lower earnings due to} \\ \text{impact of reduced} \\ \text{education } (\partial e / \partial I_1)}} < 0.$$

As displayed in the above equation, the impacts of childhood smoke exposure affect earnings through two channels. The first term captures the direct channel: lower health stock could cause lower earnings through poor physical health (which likely manifests through reduced extensive-margin labor supply), lower cognitive ability, or reduced non-cognitive abilities (both of which could manifest as either reduced labor supply or lower wages). The second term captures the indirect effect of educational attainment. Insofar as individuals have lower levels of education due to wildfire smoke exposure, this will translate to lower earnings. As part of our analysis in Section 7.2, we perform back-of-the-envelope calculations to assess how much of the earnings effect is potentially transmitted through this education channel.

The last stage of the model concerns health in later adulthood and longevity ($h_2$), which is determined as a function of childhood health ($h_1$) and later-life health inputs ($I_2$).[12] Adult health inputs are themselves based on an individual's education and earnings, $I_2 = I_2(e, \ell)$. Taken together, this gives:

$$\begin{aligned} h_2 &= h_2(h_1, I_2(e, \ell)) \\ &= h_2(h_0, I_1, I_2(e(h_0, I_1), \ell(e(h_0, I_1), h_0, I_1))). \end{aligned}$$

Thus, the impact of a childhood wildfire shock on later life health is written as:

$$\frac{\partial h_2}{\partial I_1} = \underbrace{\left[\frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial I_1}\right]}_{\substack{\text{Worse health due to} \\ \text{direct/sustained} \\ \text{effects}}} + \underbrace{\left[\frac{\partial I_2}{\partial e} \cdot \frac{\partial e}{\partial h_1} \cdot \frac{\partial h_1}{\partial I_1}\right]}_{\substack{\text{Lower investments due to} \\ \text{adverse health impact on} \\ \text{education } (e)}} + \underbrace{\left[\left(\frac{\partial I_2}{\partial \ell} \cdot \frac{\partial \ell}{\partial h_1} \cdot \frac{\partial h_1}{\partial I_1}\right)\right.}_{\substack{\text{Lower investments due to} \\ \text{adverse health impact on} \\ \text{earnings } (\ell)}} + \underbrace{\left.\left(\frac{\partial I_2}{\partial \ell} \cdot \frac{\partial \ell}{\partial e} \cdot \frac{\partial e}{\partial h_1} \cdot \frac{\partial h_1}{\partial I_1}\right)\right]}_{\substack{\text{Lower investments due to} \\ \text{impact of reduced education} \\ (\partial e / \partial I_1) \text{ on earnings } (\ell)}}$$

The first term details the direct effect of early-life health on later-life health. Insofar as reductions in early life health capital are permanent, this term will be negative. The second term relates to the indirect effect of lower educational attainment, which could increase adverse health behaviors. While the evidence on education and later-life health is mixed, in Section 7.2 we utilize estimates from the literature to assess the role of this channel. Finally, the last two terms capture the effects of wildfire smoke transmitted through the earnings channel. Lower earnings in life could result

[12] It would also be reasonable to utilize a model with many periods, where adult health and earnings are determined contemporaneously within a given period and act as inputs into the next period's outcomes. However, we abstract away from such a scenario in favor of simplicity and expositional clarity.

in lower health investments. To the extent possible, we evaluate the degree to which our mortality estimates can be explained by lower earnings (Section 7.1) or other socioeconomic factors, such as occupation and neighborhood choice (Section 9.3).

# 3 Data

To estimate the effects of wildfire smoke on long-run outcomes, we construct a longitudinal data set linking childhood measures of wildfire smoke exposure to administrative and large-scale survey data detailing adult measures of longevity, economic achievement, educational attainment, and disability in adulthood. A simplified version of this data construction process, presented in Appendix Figure A3, displays data sources and key steps used to clean and assemble our data. The sections that follow discuss these items in greater detail.

## 3.1 Wildfire Data

Our analysis utilizes wildfires from California, the state where wildfires are most severe and have been tracked for over a century. Specifically, we obtain "fire atlases" from the California Department of Forest and Fire Protection's ("CAL FIRE") Fire Perimeter Database, which is "the most complete digital record of fire history in California" (California Department of Forestry and Fire Protection, 2018). Two characteristics of this database make it ideally-suited for our analysis of the long-run effects of air pollution from wildfire smoke. First, the database includes comprehensive data on fires for nearly 100 years—and includes some coverage of fires as early as 1878. This consistent and long-running tracking of fires allows to measure differential early-life air pollution exposure for children born as early as 1930. The availability of this historical data is a vital feature for our analysis. Because we can observe children born many years ago, we are able to evaluate longer-term effects—such as longevity and mid-to-late adult outcomes—that were previously unmeasurable with existing data.[13]

Second, the CAL FIRE data is highly detailed—it includes geospatial data for each fire (i.e., the exact area that was burned), the acreage burned, and the date that each fire was discovered. These characteristics are key inputs into the pollution modeling process discussed later in this section and allow us to generate strong predictions of the cohorts who were exposed to smoke pollution. To the best of our knowledge, there is no other source that combines both the number of years and the quality of data present in the fire atlas data.[14]

[13] Specifically, most studies of the long-run effect of childhood air pollution utilize pollution monitor data (largely available starting in the mid-to-late 1960s) and leverage exogenous variation from the Clean Air Act Amendments of 1970. Because these data sources are relatively recent, papers that use them are limited to observing adult outcomes relatively early in life.

[14] Notably, the National Interagency Fire Center has detailed wildfire records, but these date back only to 1983. Additionally, the United States Forest Service maintains historical fire perimeter databases for much of the Western United States, but these these appear to be incomplete, with many lacking any information on the date of the fire, a key input into the pollution modeling analysis.

Although the Fire Perimeter Database has high-quality information regarding Californian wildfires, 42% of fires during our study period do not have exact dates when the fire occurred—only the year of the fire. To complete this missing information, we hand-gathered dates for each fire from historical newspaper archives.[15] Overall, 512 fires were missing exact dates and required hand-dating; of those fires, we were able to assign dates to 458 (88%) with high confidence. The remaining fires were treated as missing. Notably, fires that could not be confidently dated tend to be fires that are smaller and/or are in remote areas and are therefore less likely to result in harmful pollution. For purposes of our analysis, we focus on fires that burned 1,000 acres or more, which comprise approximately 89% of the total acres burned during our study period (1,232 fires in total).

Figure 1 displays statistics regarding the severity of Californian wildfires—measured by total acres burned—over time. Panel A focuses on the study period of 1930 through 1969 and demonstrates that there is substantial temporal variation in the number of burned acres: some years (1937 and 1963) there are fewer than 20,000 acres burned while others (1950) reach nearly 500,000 acres.[16] To provide further context for our sample period and to motivate the growing problem of wildfires, we also include Panel B, which displays the temporal trends of fires from 1930 through 2021. Within the figure, which has a vertical axis eight times the height of Panel A, we see two trends. First, the variation in acres burned was relatively constant from 1930 through 1990. However, there has been a marked increase in the severity of fire seasons over the last 25-30 years. In fact, the four most extreme seasons in California's recorded history have all occurred within the last five years, emphasizing the potential for these fires to have a growing impact on health and economic activity. In addition to the temporal variation in fires, Appendix Figure A4 presents the spatial distribution of fires (in 10-year bins), underscoring the fact that wildfire hazard is dispersed throughout the state.

### 3.2 Smoke Data

A central challenge of assessing the health impacts of wildfire pollution is understanding which individuals were exposed to smoke. Recent papers studying contemporaneous impacts of recent wildfire pollution—such as Miller et al. (2017) and Borgschulte et al. (2020)—utilize satellite remote sensing data to identify smoke plumes, which is available for the years 2005 through 2021. However, no such data is available during the sample period in which our cohorts were born. When the ability to directly observe pollution is absent, papers have historically assigned treatment using simple proximity to the pollution source (e.g., Clay et al., 2016) or by utilizing a combination of wind direction and proximity to the pollution source (e.g., Schlenker and Walker, 2015; Rangel and Vogl, 2019). We build on these techniques by utilizing sophisticated pollution modeling tools,

[15] Fires were located in newspaper records by searching internet archives—`newspapers.com` and `genealogybank.com`—for key terms. For instance, a typical search included typical key words, such as "acres burned", along with information gathered from the Fire Perimeter Database: the year of the fire, the name of the forest (e.g., "Mendocino National Forest"), and a prominent geological feature near to the fire (e.g., "Clear Lake").

[16] A potential concern with historical data is that recorded fires could increase over time solely due to improvements in tracking them. However, we note that there is no statistically detectable trend ($p = 0.58$) in the number of acres burned over the sample period, alleviating this concern.

joining recent environmental economics papers such as Sullivan (2017) and Grainger and Ruangmas (2018) who integrate these models into their empirical approach. Specifically, we utilize the National Oceanic and Atmospheric Administration's ("NOAA") HYSPLIT model to predict dispersion of wildfire smoke (Stein et al., 2015). HYSPLIT, which is used by the U.S. Forest Service to predict wildfire pollution, utilizes the location, date, and size of the fire (in acres) as inputs and combines them with gridded four-dimensional meteorological data and fuel composition data obtained from the U.S. Forest Service's BlueSky modeling framework (Larkin et al., 2009) to predict smoke plumes from each fire that compare favorable to satellite observations. In addition to the satellite comparisons performed by existing research, we also validate the model-generated smoke plumes by assessing their impact on ground-level pollution. This analysis, which is detailed in Section 4, finds that modeled smoke exposure is highly predictive of elevated pollution in a modern (2004-2019) context.

Our use of pollution modeling provides several advantages over the proximity or wind direction methods. First, HYSPLIT incorporates the amount of fuel burned, which is a function of both (a) acres burned and (b) the underlying fuel source (i.e., the density of forest or brush in the area of the fire). Given that the sizes of fires in our sample range from 1,000 acres to over 200,000 acres and occur everywhere from brush land to densely wooded forests, this combination is critical in accurately assessing the amount of air pollution from wildfires. Second, pollution modeling utilizes wind dynamics in a more sophisticated manner than is typically observed in existing methods by further incorporating wind speed and wind direction outside of the immediate vicinity of the fire. Wind speed is an important consideration for fires because calm winds can result in smoke patterns that diffuse around the origin point of the fire, while strong winds can send fire pollution for hundreds—if not thousands—of miles in a specific direction. Because wildfire smoke can travel so far from its origin, inclusion of wind dynamics that are not in the immediate vicinity of the fire are also necessary to correctly assign exposure to wildfire smoke. These considerations are illustrated by Appendix Figure A5, which displays modeled smoke output under three different scenarios with varying fire size and wind intensity. In each scenario, the smoke plumes differ meaningfully, underscoring the importance of incorporating fire size and rich wind dynamics into the treatment definition.

While utilizing HYSPLIT provides several advantages over existing treatment definitions, there are two limitations to the smoke plume output. First, HYSPLIT generates smoke plumes as if the entire acreage is burned within first twenty-four hours from the start of the fire. Because wildfires often last several days, we generate smoke plumes for three separate dates: the date in which the fire began and two subsequent days.[17] Second, the model allows use of meteorological data only back to 1948, which is roughly in the middle of our sample period. To compensate for this, we match each fire date from *before* 1948 with a date *on or after* 1948. This matching process is done by finding the date with the smallest Euclidean distance across several different measures of wind direction and intensity at the ground level as well as multiple levels of the atmosphere. The details

[17] Areas are considered to be treated if they spatially intersect any of the plumes generated by this procedure.

of this matching process are covered in more detail in Appendix Section A.1.

The identifying variation for our analysis is displayed in Figure 2, which details measures of smoke exposure across birth cohorts and across space. Panel A demonstrates the degree to which birth cohorts are exposed to wildfire smoke during the period from *in utero* through age 4 ("IU-Age 4").[18] As further discussed in Section 5, we focus on this particular age range because that is when children appear to be most sensitive to wildfire shocks. Figure 2 Panel A displays the mean, interquartile, and interventile ranges of *in utero* through age 4 smoke by year-of-birth cohort, along with the mean and interquartile range ("IQR") for the entire sample. The figure illustrates that the amount of exposure varies substantially across California birth cohorts, with peak exposure for cohorts born in the 1950s and lowest exposure for 1930s cohorts.

Even within cohorts, there is substantial range in exposure, as demonstrated by the meaningful differences between the most- and least-intensely treated individuals. This within-cohort variation is more fully illustrated by Panel B, which demonstrates how smoke exposure—defined as the percentage of months in a given period with any smoke coverage—varies over geography in ten-year bins. As displayed in the figure, there are clear geographic differences in exposure to smoke pollution. However, these differences do not come from the same heavily and lightly treated areas for every birth cohort. For instance, in some years, Northern California is more intensely exposed, while in other years Southern California is subjected to higher smoke pollution. As discussed above, this is due in part to fluctuations in fire timing and severity. It is also a function of wind speed and direction, factors that vary meaningfully, as demonstrated by Appendix Figure A6. As discussed further in the sections to follow, these environmental determinants of smoke exposure provide plausibly exogenous variation in our treatment variable.

### 3.3 Outcome Data and Linkages using Restricted Census Data

To study the long-run effects of childhood exposure to wildfire pollution, we link smoke exposure in early life to three restricted-use datasets from the United States Census Bureau: the Social Security Administration's ("SSA") Numident file, the 2000 Decennial Census Long Form (a 1-in-6 sample of households) and the 2005-2019 American Community Surveys ("ACS").[19] There are three primary features about these data make them ideal for our analysis. First, the Numident file—which is an administrative database of all individuals who have ever been assigned a Social Security number—includes both a person's city and state of birth, as well as their exact date of birth. These characteristics allow us to assign measures of smoke exposure to individuals in a very precise man-

[18] While we do not observe actual gestational lengths in our data, an individual is assumed to have *in utero* exposure if there is a smoke plume present within the nine months before their birth. Further, exposure "through age 4" indicates smoke exposure at any time prior to a child's 5th birthday.

[19] The SSA Numident file that we obtain through Census Bureau differs from other versions of the Numident database that are available publicly, such as the Berkeley Unified Numident Mortality Database (Goldstein et al., 2021), primarily in the larger number of cohorts covered by the SSA/Census Numident and the ability to link to other data sets (discussed further below). Nonetheless, certain descriptive analyses in this paper will use the Berkeley Unified Numident Mortality Database (henceforth referred to as the "Public-version Numident") to minimize the amount of restricted Census data that is subject to disclosure.

ner.[20] Second, in addition to detailed information about birth, the Numident file includes mortality information for all individuals in the database. This mortality data is particularly comprehensive for deaths occurring after the mid-1970s (Chetty et al., 2016; Finlay and Genadek, 2021), which we discuss further below. Third, the restricted datasets can be linked together by using a unique, individual-level identifier (a Protected Identification Key, which is essentially a scrambled Social Security number). This linkage allows us to apply our childhood exposure measures—assigned using an individual's birth place and birth date information from the Numident data—to adult survey outcomes in the 2000 Census, ACS, and other Census products.

We focus primarily on two sets of outcomes in our analysis. First, we evaluate the effects of early-life smoke pollution on longevity. Our measure of longevity is conditional upon surviving past childhood, because individuals need to obtain a Social Security number prior to inclusion in the Numident data. Within our sample, individuals obtain their Social Security number at 14.4 years of age, on average.[21] We evaluate mortality by age 55 as our primary outcome, as that is the oldest age at which we observe all cohorts in our data, and we also evaluate mortality at ages 30 through 80 to calculate a measure of life years lost due to wildfire smoke. The second set of dependent variables is taken from the 2000 Census Long Form and American Community Surveys. Because these surveys provide such a rich set of measures—and because childhood circumstances can affect such a wide range of adult outcomes, as illustrated in Section 2—we organize survey measures into three distinct groups: (1) economic achievement, (2) educational attainment, and (3) Census-defined disability measures. For each group, we create indices of standardized outcomes, following Kling et al. (2007), to mitigate concerns about multiple hypothesis testing. To construct these indices, individual component variables are converted to z-scores by subtracting the mean value across cohorts and dividing by the standard deviation of the outcome. Certain variables—poverty status, social safety net use, and disability variables—are then reverse coded as necessary, so that higher values represent "good" outcomes. Finally, each variable is combined into an index by taking a simple mean of these standardized values, where each variable is assigned an equal weight in the index. While we focus on these index outcomes as our main results, we will also present effects for individual index components.

We limit all samples, with the exception of our longevity sample, to only include individuals who are younger than age 60 at the time of the survey in order to isolate the impact of wildfires on individuals during their peak earning years. Additionally, we only include individuals who do

[20] The ability to observe *city* of birth is a substantial improvement over most data sources, which only provide state or country of birth—if they provide any birth place information at all. Having city-level data—rather than only county-level data—is particularly helpful in our setting of California because many counties in California have very large land areas. (The *median* county is nearly the size of Delaware and the largest is almost the size as West Virginia.) Thus, having city of birth detail considerably improves the accuracy of treatment assignment. Additionally, the ability to observe exact date of birth is also a substantial improvement over public data sources, which are typically limited to birth year.

[21] The date of Social Security number ("SSN") receipt is proxied by the "initial cycle date" in Numident. It is worth noting that the SSA's "Enumeration at Birth" program, which allows parents to complete SSN applications for newborns as part of the birth registration process, did not start until 1987. Accordingly, individuals in our sample typically obtained their SSN in late childhood or adolescence.

not have any imputed or missing outcomes for all index components to ensure consistency across analyses. Finally, we restrict our sample to individuals who had obtained their Social Security Number ("SSN") by age 19, which includes 95.6% of individuals in our sampling frame, to ensure that any deaths during adulthood can be captured in the Numident death file.[22] After applying these restrictions, the number of observations in our core samples include approximately 6,656,000 individuals in the longevity/mortality sample and 897,000 person-years underlying our analysis of economic, education, and disability indices (and their sub-components). The summary statistics for mortality and index component outcomes are displayed in Appendix Table A1.

While the SSA Numident file provides key information necessary for our analysis, there are two challenges to using the data. First, place of birth data is generated by information that individuals enter when completing the application for a Social Security number. These responses are stored as non-standardized string variables, which include spelling errors, transcription errors, use of abbreviations, and/or inclusion of places that do not qualify as cities. For example, an individual born in Los Angeles, California might write their city of birth as "Los Angeles," "Los Angels," "Westwood LA" (a neighborhood in Los Angeles), or simply "LA." In order to standardize these responses, we limited our sample to all individuals who reported their birth state as California and then hand-matched common responses to populated places detailed in the United States Geological Survey's Geographic Names Information System ("GNIS") file. The GNIS file includes the county, longitude, and latitude of each location.[23] The remaining observations were matched using probabilistic algorithms, which are further discussed in Appendix Section A.2. Overall, we matched 99.5% of individuals to a GNIS location, where 99.1% of these matches either had a perfect text match or were hand-matched.

Second, there are challenges when utilizing Numident to evaluate longevity, the main outcome of our analysis. Because Numident began tracking deaths in the early 1960s—and not comprehensively until the mid-1970s (Finlay and Genadek, 2021)—rates of death-by-age are understated, particularly for our earliest cohorts. To correct for this, we constructed adjusted death rates that scale death rates per the Numident by death rates from the Social Security Administration's Cohort Life Tables ("SSA Cohort Life Tables"), a process that is discussed in greater detail in Appendix Section A.3. As demonstrated by Appendix Figure A8, the adjusted death rates provide a much better approximation to expected cohort death rates than the unadjusted values. Additionally, adjusted death rates for earlier cohorts (those born in the 1930s and 1940s) are more comparable with rates for cohorts born from 1950-1960, further suggesting that adjusted rates are more appropriate as the primary outcome. Nonetheless, we demonstrate in our Section 6 analysis that results are robust to using unadjusted rates.

[22] We include this sample restriction because individuals who obtain Social Security numbers later in life may be subject to survival bias.

[23] Longitude and latitude relate to the centroid of each place. A potential concern with utilizing city centroids is that the location of these centroids could drift over time as city boundaries change. However, as demonstrated by Appendix Figure A7 and discussed in the accompanying notes, this is likely not an issue because boundaries are quite stable over time.

### 3.4 Other Data

In addition to our longitudinal data linking early-life wildfire smoke exposure to long-run mortality and survey outcomes, we also utilize several data sources for the purpose of supplemental and validation analyses, which we briefly outline here.

**Restricted Decennial Census Full Count Data, 2000 and 2010**. These data are drawn from the "short form" questionnaire that is received by all households in a given Census year and accordingly cover nearly the entire U.S. population. The Full-Count Census files are then linked to early-life smoke exposure measures in the same manner discussed in the previous subsection. We utilize the linked data to perform supplemental analysis of neighborhood characteristics as well as certain family formation and long-run migration analyses, all of which are discussed in Section 9.

**Environmental Protection Agency ("EPA") Monitoring Data, 2004-2019**. The EPA collects daily monitor readings for a variety of pollutants. In order to relate this monitoring data to smoke exposure, we generate modeled smoke plumes using 2004-2019 wildfire data from CAL FIRE and the techniques discussed in Section 3.2. These data are incorporated into the analysis in Section 4 which demonstrates that the modeled smoke measures are highly predictive of ground-level pollution.

**Control Variable and Heterogeneity Data**. We also utilize a variety of other variables—namely county-level aggregates, historical hospital data, and climate measures—to include in our robustness and in heterogeneity analyses (Section 8 and Section 9, respectively). The data sources from which these variables were derived, and our method for incorporating them into our main analysis data set, are discussed further in Appendix Section A.4.

## 4 Validation of Smoke Measures

Because modeled smoke exposure discussed in Section 3.2 will form the basis for our treatment variable, we show in this section that it strongly predicts ground-level air pollution in a modern context.[24] In order to do so, we estimate the following distributed leads-and-lags regression model:

$$Y_{it} = \gamma_{y(t),i} + \eta_{m(t),i} + \sum_{j=-4}^{4} \beta_j AnySmoke_{i,t-j} + \varepsilon_{it}, \tag{1}$$

where the unit of analysis is a monitor $i$ with associated pollution reading observed in year-month $t$. To control for time-varying and seasonality-specific trends for each monitor, monitor-by-year fixed effects ($\gamma_{y(t),i}$) and monitor-by-calendar-month fixed effects ($\eta_{m(t),i}$) are included in the regression.

The treatment variable, $AnySmoke_{i,t-j}$, is equal to one a monitor was exposed to smoke during period $t-j$. Specifically, we consider a monitor to be "exposed" if the modeled smoke plume spatially intersects with the monitor's geocoordinates. We include leads and lags of the treatment variable to (1) account for serial correlation in smoke exposure and (2) trace out the effect of smoke

[24] We do not validate our measure in a *historical* context because monitor data does not exist for long-ago periods.

exposure on pollution over time. The coefficients of interest on these leads and lags, $\beta_j$ $\{j \in -4...4\}$, represent the marginal effect of smoke coverage during relative month $j$ on a monthly pollution measure during month $t$.

The main results of this analysis are displayed in Figure 3, which shows the impact of predicted wildfire smoke plumes on *mean* daily $PM_{2.5}$ and *max* daily $PM_{2.5}$ in Panels A and B, respectively. Within the figure, the estimated impact of smoke coverage during relative month $j$ is presented with 95% confidence intervals. The estimated impact on pollution during the month in which the fire occurs ($\beta_0$) is presented in the upper-right-hand corner of each panel, along with the associated standard error and baseline mean.

We find that our modeled smoke exposure predicts meaningful increases in ground-level $PM_{2.5}$ during the month in which the fire occurs: mean daily pollution increases by 1.26 $\mu g/m^3$ (13% over the mean) and max daily pollution by 3.58$\mu g/m^3$ (18% over the mean). These effects decay over the two months after the start of the fire, as fine particulates remain in the air for weeks at a time.[25] In addition to the impacts on $PM_{2.5}$, smoke exposure has an impact on a range of other pollutants. As displayed in Appendix Figure A9, we find increases in mean and maximum for larger particulate matter ($PM_{10}$) of 9% and 15%, respectively, alongside 2-5% increases over the mean for Carbon Monoxide (CO), Ozone ($O_3$), and Nitrogen Dioxide ($NO_2$). Taken together, these results support the assertion that our modeled wildfire smoke measure is an effective predictor of pollution.

## 5 Methodology

To evaluate the long-run effects of early childhood wildfire smoke exposure, we estimate the following cohort distributed lags and leads model:

$$Y_{cba} = \mu_c + \lambda_{y(b),r(c)} + \alpha_{m(b)} + \psi_{y(b),a} + \sum_{k=-5}^{5} \delta_k Smoke_{c,b+k} + \varepsilon_{cba}, \tag{2}$$

where the unit of analysis is a birth cohort born in a city $c$ in year-month $b$ and observed at survey year $a$. The treatment variable, $Smoke_{c,b+k}$, and its associated leads and lags detail the number of months with any smoke exposure at age $k$, where a month of smoke exposure is defined as any spatial intersection between a smoke plume and a child's city of birth during that month (see Appendix Figure A10 for an example of how treatment is defined). The coefficients of interest on these treatment variables, $\delta_k$ $\{k \in -5...5\}$, provide the marginal effect of an additional smoke month during age $k$. The $\delta$ coefficients for the *in utero* period through age 5 ($\delta_{IU} \dots \delta_5$) allow us to trace out the dynamic effects of additional smoke exposure at various ages. Meanwhile, the coefficients for the years leading up to conception ($\delta_{-5} \dots \delta_{-1}$) measure the impact of smoke

[25] The slow decay of the effects is also likely due to the fact that some fires, particularly in recent years, can last for a very long time (>1 month). However, because wind direction is autocorrelated over short periods, our modeled smoke measure is still able to capture these lagged pollution effects, even though we are unable to explicitly incorporate the duration of the fire into our model.

exposure that occurs before a child could be conceived. Accordingly, these estimates collectively serve as a placebo test, since no effect is expected from air pollution exposure during these periods.

We include city-of-birth fixed effects ($\mu_c$) to control for time-invariant factors related to individuals born in the same city.[26] Additionally, we include year-of-birth $\times$ region-of-birth fixed effects ($\lambda_{y(b),r(c)}$) where region of birth is defined in our baseline specification as Northern or Southern California.[27] These fixed effects capture time-varying differences by regions of California that may differ in terms of smoke exposure, climate, and economic activity. Further, because both birth outcomes (Buckles and Hungerman, 2013) and wildfires are highly seasonal, we include calendar-month-of-birth fixed effects, $\alpha_{m(b)}$.[28] Finally, when utilizing non-mortality outcomes as the dependent variable, we include year-of-birth by age-at-survey fixed effects ($\psi_{y(b),a}$) to ensure comparisons are made between individuals who were both born and surveyed in the same pair of years. Standard errors for estimates from this equation are clustered at the county level.[29]

In addition to our model detailed in Equation 2, we also estimate an equation that collapses the effects of childhood wildfire smoke exposure into a single variable:

$$Y_{cba} = \mu_c + \lambda_{y(b),r(c)} + \alpha_{m(b)} + \psi_{y(b),a} + \delta \cdot Smoke_c^{\text{IU-Age 4}} + \varepsilon_{cba}. \tag{3}$$

The treatment variable, $Smoke_c^{\text{IU-Age 4}}$ is a single continuous variable equal to the number of months with any smoke exposure from the *in utero* ("IU") period through age 4.[30] For notational convenience, estimates from this equation are referred to as our "summary estimates," "IU-Age 4 estimates," or "$\delta_{\text{IU-Age 4}}$" throughout the remainder of this paper.

We focus on the *in utero* through age 4 period primarily for two reasons. First, as discussed extensively in Currie and Almond (2011) and Almond et al. (2018), the years prior to age 5 are a particularly important time of development and shocks during that period have been shown to have particularly meaningful later-life impacts. Second, as will be demonstrated by the results in Sections 6 and 7, the age-of-exposure effects largely dissipate by age 5 (consistent with the existing literature) and thus we focus on the impact of smoke exposure that occurs before that age.

Our research design relies on the assumption that the intensity of wildfire smoke exposure during the first years of life are uncorrelated with other factors, conditional on our fixed effect controls. Because we control for time-invariant factors specific to a given place of birth (via our city-of-birth fixed effects, $\mu_c$) as well as region-specific, time-varying factors (via $\lambda_{y(b),r(c)}$, our vector of year-of-birth $\times$ region-of-birth fixed effects), potentially confounding factors would need to

[26] Cities with fewer than 100 births are grouped together with their nearest neighbors until each combined city group has at least 100 associated births during the sample period.

[27] The counties grouped into either Northern or Southern California a displayed in Appendix Figure A11, Panel A. In our robustness checks in Section 8, we instead utilize State Economic Areas—precursors to modern-day commuting zones—and find similar results. These areas are illustrated in Appendix Figure A11, Panel B.

[28] Seasonality in wildfires is demonstrated Appendix Figure A12, which shows that 90.3% of acres are burned during the months of July through November.

[29] This is true for all other regression estimates presented in this paper, unless noted otherwise.

[30] Because we do not observe actual gestation period in our data, the *in utero* period is assumed to begin nine months prior to birth.

be time-varying at a sub-regional (e.g., city or county) level. The leads in our distributed leads and lags model ($\delta_k$ for $k \in \{-5...-1\}$ in Equation 2) provide a useful diagnostic test as to whether or not such confounding issues exist. If such confounders exist, then these coefficients will likely be economically and statistically different from zero. Additionally, in Section 8, we consider different fixed effect and control variable configurations to address the possibility of unobserved confounders that are also conditionally correlated with wildfire smoke exposure and find that results are robust to these alternate specifications.

# 6 Long-Run Mortality Effects

The first outcome we estimate is the long-run impact of childhood wildfire smoke exposure on longevity, specifically survival to age 55, which is the oldest age that we observe all of the cohorts in our sample. (Recall that all analyses are conditional on surviving past childhood, as it is necessary to obtain a Social Security number to be included in the data.) The results of this analysis are displayed in Panel A of Figure 4 where each point represents a value of $\delta_k$ from Equation 2 along with associated 95% confidence intervals. Each point estimate represents the marginal effect of an additional smoke month during age $k$, where the *in utero* age ($k = \text{IU}$) represents the nine months preceding birth and negative ages ($k \leq -1$) represent years prior to the *in utero* period. Additionally, our measure summarizing the marginal impact of an additional month with exposure during the *in utero* through age 4 period (described by Equation 3) is presented in the upper-right-hand corner along with the associated standard error.

The results in Figure 4 demonstrate a clear effect of air pollution from wildfires during early childhood: our summary estimates show that an additional month of smoke exposure during the *in utero* through age 4 period ($\delta_{\text{IU-Age4}}$) results in 0.34 additional deaths before age 55 per 1,000 individuals. Moving from the 25$^{\text{th}}$ to the 75$^{\text{th}}$ percentile—an increase of 5 months—results in 1.7 additional deaths before age 55 per 1,000 (for conciseness, the effect of moving from the 25$^{\text{th}}$ to the 75$^{\text{th}}$ percentile will henceforth be denoted as $p_{25} \rightarrow p_{75}$). This effect is a 1.8% increase over the age 55 cumulative death rate within our sample. The leads and lags coefficients, which show the age-specific exposure effects, demonstrate when childhood smoke exposure matters the most. We find that the effects of exposure are largest around age 2 and have dissipated by age 5.[31] We do not, however, find any effects for our placebo estimates of smoke exposure before conception (relative periods -5 through -1), supporting the assertions that exposed areas are not differentially trending prior to exposure or subject to changing birth composition in anticipation of future wildfire smoke shocks (see Section 8 for further discussion).

[31] This pattern—specifically that the long-run impacts of air pollution exposure fades by age 5—is consistent with research on long-run *wage* impacts by Colmer et al. (2022), who find that the effects of the 1990 CAAA on earned persist through age 4 and decrease dramatically thereafter.

There are some concerns that the attenuation of effects by age 5 could be due to measurement error, since treatment is assigned based on city of birth and thus may incorrectly assigned due to migration. We bound this measurement error using migration data from the 1940 Complete-Count Census (Ruggles et al., 2021) and find that it does not qualitatively affect the results. See Section 8 for details.

While studies of the *contemporaneous* effects of air pollution find the strongest impacts when children are *in utero* or infants, our age-specific effects imply that the most sensitive period for severe, transitory wildfire smoke shocks occurs when children are slightly older (ages 1-2). This could be for two primary reasons. First, slightly older children are likely exposed to much higher levels of wildfire smoke because they spend more time outside and have higher lung capacity. Time outside matters a great deal because smoke exposure is substantially higher outdoors—for instance, Burke et al. (2021) find that outdoor $PM_{2.5}$ can be approximately 10 to 20 times higher than indoor measures during periods of heavy wildfire smoke. Further, because wildfire smoke tends to be transitory, this ratio is much higher than for persistent pollution sources that have time to infiltrate homes (Liang et al., 2021). Accordingly, it is reasonable that children who are still young enough to be vulnerable—but are old enough to spend more time outdoors during periods of elevated air pollution—have more pronounced responses to wildfire smoke. Second, because our result is conditional on surviving past childhood, it excludes deaths that occur very early in life. Accordingly, insofar as individuals die before obtaining a social security number—something which occurs at age 14, on average, in our sample—they will be excluded from this mortality calculation. We consider this by estimating cohort-level infant mortality in Appendix Section B.1 and adding this mortality into our estimates. While these added deaths increase the response to air pollution in early life, our effects are still most pronounced for slightly older children.

We further explore the effects of wildfire air pollution on cumulative mortality across a variety of ages in Figure 4, Panel B. Specifically, we estimate summary coefficients ($\delta_{\text{IU-Age4}}$) for regressions where the outcome is cumulative mortality by age $a$ (where $a \in \{30, 35, ...80\}$). The results of these separate regressions are displayed in Panel B of Figure 4 along with 95% confidence intervals. We find that small effects manifest relatively early (age 30) and return to zero by age 80.[32] Note that age effects for *any* estimate of cumulative mortality must mechanically "fade" to zero as all members of the counterfactual group eventually die of old age. In our context, it appears that wildfire pollution shocks do not appreciably affect mortality after age 80, although the estimates are highly imprecise. To aggregate these mortality effects, we calculate the life years lost between ages 30 to 80 and find that an additional month of wildfire smoke exposure results in 9.2 life years lost per 1,000 persons ($p_{25} \rightarrow p_{75} = 46$ per 1,000).[33] As discussed in Section 3.3, our mortality data are adjusted to better match aggregate death patterns in our sample period. We re-estimate the analyses detailed above using the *unadjusted* rates and obtain similar estimates—8.1 life years lost per additional smoke month compared to 9.2 life years when using adjusted death rates. See Appendix Figure A13 for more detail.

To further contextualize our mortality estimates, we compare the cohort-level life years lost due to smoke-induced infant mortality to the reduction in life years attributable to decreased adult longevity. As detailed in Appendix Section B.1, we calculate the effects of wildfire pollution on infant mortality using California vital statistics data and a cohort distributed leads-and-leads model

[32] Tabular data on point estimates and associated standard errors are presented in Appendix Table A2, Panel A.
[33] Standard errors for this value were obtained using bootstrap (100 iterations).

in a similar spirit to the one specified in Equation 2. From our estimates, we calculate that moving from the 25th to 75th percentile of smoke exposure results in 11.7 life years lost per 1,000 births due to infant mortality, as compared a reduction of 46 years per 1,000 due to shorter adult life spans, as shown above. Decreased longevity results in a larger number of life years primarily because the critical period where wildfire smoke can reduce adult longevity is substantially longer than the period where infants are at high risk of death from wildfire smoke. That is, there are simply more opportunities for air pollution from fires to reduce adult longevity. Given that reductions in infant mortality are a staple of cost-benefit analyses for environmental policy, our finding that early-life air pollution has even larger effects through the channel of reduced adult longevity underscores the importance of incorporating long-run effects of early-life harms into policy calculations.

# 7 Long-Run Effects on Economic, Educational, and Disability Outcomes

In this section, we build on the literature that examines the long-run effects of childhood air pollution by exploring economic, educational, and disability outcomes at older ages (35-59) than have previously been considered. The results of our analyses are displayed in Figure 5. Within the figure, each panel presents an event study for the main index (to the left) as well as the summary coefficients (from Equation 3) for the index and each of its components (to the right). As discussed in Section 3.3, these variables have been standardized (z-scored) so that their mean is 0 and standard deviation is equal to 1. This procedure allows us to combine many results into a single index. Further, we scale these coefficients such that the estimates presented can be interpreted as a percent of a standard deviation (henceforth, "%-SD"). In the following sub-sections, we discuss each index and provide more detail regarding the index components.[34]

## 7.1 Economic Achievement

The results relating to economic achievement are presented in Panel A of Figure 5. Within the figure, we observe a clear decrease in economic achievement in later life, with effects largely persisting through age 4 and subsiding by age 5, a pattern that is qualitatively similar to the mortality results discussed above. Aggregating across the *in utero* through age 4 period, we find that the average impact of an additional month of smoke exposure in is equal to 0.29%-SD ($p_{25} \rightarrow p_{75}$ =1.45%). Upon examining the index components, we find that these effects are consistent across outcomes: there are decreases in income, extensive-margin employment, home ownership, and home value. Further, there are *increases* in the likelihood of being in poverty and requiring use of the social safety net.[35]

Our "headline" economic achievement outcome—i.e., the outcome that typifies our results—is the decrease in earned income (defined as wages plus self-employment income). As indicated by

[34] Tabular data on point estimates and associated standard errors for this figure are presented in Appendix Table A2, Panel B.

[35] Poverty and social safety net variables have been reverse coded so that lower values indicate worse outcomes.

Panel A of Appendix Figure A17, we find that an additional month with smoke exposure reduces age 35-59 earned income by \$178 per year ($p_{25} \rightarrow p_{75}$ = \$890), relative to a mean of \$44,087.[36] We calculate the cumulative impact of these lost earnings and discount them back to birth using a 3% discount rate. After incorporating the evolution of our treatment effects over the life cycle—which are discussed in detail in Section 9.1—we find the $p_{25} \rightarrow p_{75}$ effect of childhood wildfire pollution exposure reduces discounted earnings over the age ranges in our sample by \$4,710. This suggests that wildfire pollution can have substantial long-run effects on earnings and economic productivity.

To provide context for the magnitude of these effects, we combine our estimates above with those from Figure 3 to calculate the relationship between changes in fine particulate matter and later-life earnings. Our findings imply that a 1 $\mu g/m^3$ increase in *annual* $PM_{2.5}$ results in a 2.5% decrease in earned income. This result is close to the most recent estimates of the long-run impact of air pollution on early adult wages by Colmer et al. (2022), who find that a 1 $\mu g/m^3$ decrease in annual $PM_{2.5}$ results in a 2.8% to 4.3% increase in wages. We further note that our age-of-exposure effects align with those found in Colmer et al. (2022), who also find the long-run impacts of childhood air pollution exposure fade dramatically after age 4. Nonetheless, while these similarities are reassuring, there are many reasons why we might expect the magnitudes to differ, which we discuss in further detail in Appendix Section B.2.

Next, we explore the potential causes and consequences of the long-run impacts on income. First, we consider the degree to which our earnings effects appear to be driven by decreases in labor force participation ($p_{25} \rightarrow p_{75}$ = -0.45 percentage points). To do so, we perform a back-of-the-envelope calculation which suggests that the effects on extensive-margin labor supply accounts for roughly 27% of the effect on earned income.[37] Second, we consider the degree to which the decrease in economic activity, and specifically earnings, may *contribute* to the reduction in longevity by performing an additional back-of-the-envelope calculation. We utilize data from Chetty et al. (2016) to calculate the association between income and life expectancy, and we find that a \$1,000 increase in income is associated with an increase in life expectancy of 0.030 years.[38] Accordingly, our \$890 decrease in earnings may be associated with a reduction of 0.027 life years per person, which suggests that the reduction in earnings could explain as much as 59% of the reduction in life expectancy that we find.[39]

[36] These dollar values, along with all dollar values for the remainder of the paper, are in 2010 dollars.

[37] Specifically, within our sample, a 1 percentage point increase in labor force participation is associated with a \$530 increase in earned income. Using this relationship, we can roughly calculate the fraction of the earned income effect "explained" by labor force participation effects as follows:

$$\frac{0.45\text{pp} \times \$530}{\$890} \approx 26.8\%$$

[38] We exclude the top income percentile when calculating this slope, as that is consistent with our winsorized income measures.

[39] We view this calculation as placing an *upper bound* on the degree to which early life shocks affect mortality through the income channel. This is because the relationship from Chetty et al. (2016) is not causal, and so the degree to which our economic results "explain" life expectancy reductions may be overstated. This overstatement will depend

In addition to highlighting our earnings outcome, we also briefly discuss the degree to which early childhood wildfire smoke exposure increases use of the social safety net, which is defined as any use of the Supplemental Nutrition Assistance Program, Medicaid, Supplemental Security Income, or public assistance income. We find that an additional month with any exposure increases social safety net use by 0.09 percentage points (0.66% over the mean; $p_{25} \rightarrow p_{75} = 0.45$).[40] To quantify the effect of this increased use, we utilize the Current Population Survey provided by IPUMS (Flood et al., 2021), which quantifies the household- or person-level cost of each of these programs. Using these data, we calculate that a person using at least one of these programs increases government expenditures by $7,958 annually, on average. Multiplying this cost by our estimates and discounting back to birth, we find that moving from the 25$^{th}$ to the 75$^{th}$ percentile of childhood smoke exposure results in an additional $229 of government expenditures per person.[41]

### 7.2 Educational Attainment

We next consider the effect of smoke exposure on educational attainment, which is displayed in Figure 5, Panel B. As with previous outcomes, there is a pronounced effect on educational attainment: an additional month with smoke exposure decreases the educational attainment index by 0.27%-SD ($p_{25} \rightarrow p_{75} = 1.35\%$). Moreover, there is a similar pattern in the outcome dynamics, with effects peaking at ages 1-2 and effectively trending back to zero by age 5.

We begin our discussion of index components with our "headline" outcome: high school completion. As displayed in Appendix Figure A17, Panel B, we find that a marginal smoke month results in a 0.08 percentage point decrease in high school completion ($p_{25} \rightarrow p_{75} = 0.40$pp). These effects generally persist across levels of higher education attainment, with estimated reductions in 4-year college ($p_{25} \rightarrow p_{75} = 0.50$pp) and graduate degree completion ($p_{25} \rightarrow p_{75} = 0.30$pp).

When estimating the effects on years of schooling, rather than degree completion, we estimate a $p_{25} \rightarrow p_{75}$ reduction of roughly 0.05 years of education. Using this estimate, we can conduct a back-of-the-envelope exercise to assess how much of the earnings effect (from Section 7.1) could potentially be explained by this impact on education. Specifically, we utilize a 10% of return per year of schooling (Card, 1999), which translates into a back-of-the-envelope decrease in earnings of 0.5%, or roughly 25% of the earnings effect.[42] This suggests that early-life pollution shocks may be transmitted to earnings primarily via mechanisms other than education, such as shocks to underlying ability, non-cognitive skills, later-life cognitive decline, and/or physical health.

It is also useful to consider how much of the increases in mortality, discussed in Section 6,

on the degree to which reduced health is driving lower income (rather than vice versa). Nonetheless, in the absence of causal estimates of income on life expectancy, we view bounding the income-mortality mechanism in this way to be a useful exercise.

[40] See Appendix Figure A18, Panel A, for event-study graph.

[41] Specifically, we calculate the average annual cost as $\$7,958 \times 0.45\text{pp} \approx 36$. We then discount this value back to birth using a 3% discount rate: $\sum_{a=35}^{59}[1.03^{-a} \times 36] = \$229$

[42] The wage effects found in the previous subsection translate to 2% of wages. This approximation of the education-earnings pathway is consistent with Isen et al. (2017b) who calculate that educational attainment explains approximately 20% of the earnings increases from the Clean Air Act Amendments of 1970.

could be explained by decreases in education. To do so, we utilize estimates from Clark and Royer (2013) and Meghir et al. (2018)—two papers that leverage substantial, sharp variation in years of schooling and high-quality administrative data. While these studies do not find statistically significant decreases in mortality due to higher education, they able to rule out meaningful effects. Applying their estimated bounds on the education-mortality relationship, we calculate that the estimated decreases in education could explain no more than 30% to 51% of the mortality effects that we find.[43]

### 7.3 Disability

The last set of "index" outcomes we evaluate includes all disability measures available in the American Community Surveys and 2000 Decennial Census. The leads-and-lags coefficients for our disability index—displayed in Figure 5, Panel C—show that early-life wildfire smoke exposure causes individuals to have worse later-life disability outcomes, but that these effects are small and statistically imprecise. (Recall that all disability outcomes are reverse coded so that lower values indicate worse outcomes.) However, as displayed in the summary components on the right-hand side of Panel C, the combined effect for the *in utero* through age 4 period achieves statistical significance ($t = -2.13$) with a 0.16%-SD reduction ($p_{25} \rightarrow p_{75} = 0.80\%$). In addition to our disability index, another potentially useful measure of adult health is whether an individual reports *any* disability. We find that an additional smoke month results in a 0.06 percentage point reduction ($t = 1.66$) in individuals reporting at least one disability, or a 0.4% increase over the mean. However, because of the noisiness in the event-study estimates, these results are merely suggestive.

When evaluating individual index components, we find that only one outcome—Difficulty Remembering, Concentrating, or Making Decisions (henceforth, "cognitive difficulty")—achieves statistical significance at the 95% level. The leads-and-lags graph for cognitive difficulty, displayed in Panel C of Appendix Figure A17, includes a clearer and more convincing pattern than the overall index, with effects following a similar trajectory as other outcomes. The summary coefficient, $\delta_{\text{IU-Age4}}$, implies a marginal effect of 0.06 percentage points ($p_{25} \rightarrow p_{75} = 0.30\text{pp}$). The finding that childhood air pollution exposure causes adult cognitive difficulty is consistent with the medical

[43] Clark and Royer (2013) rule out reductions in mortality larger than 5% from a 0.315 to 0.464 increase in years of education. We apply this bound to our estimates to calculate the upper bound of mortality than can be explained by education as:

$$\frac{\frac{0.05}{0.315} \times 5\%}{1.8\%} = 30\% \qquad \frac{\frac{0.05}{0.464} \times 5\%}{1.8\%} = 44\%.$$

Where the numerator in each equation scales the years-of-education effect that we find (0.05) by the 5% bound in the Clark and Royer (2013) estimates. This effect is then divided by the percentage reduction in age 55 mortality that we discuss in Section 6. Alternatively, we can use estimates from Meghir et al. (2018), who find than a 0.25 increase in years of education results in at most 1.0-1.4 *month* increase in life expectancy (0.083-0.117 years):

$$\frac{\frac{0.05}{0.25} \times 0.083}{0.046} = 51\% \qquad \frac{\frac{0.05}{0.25} \times 0.117}{0.046} = 36\%.$$

Where the numerator in each equation scales the years-of-education effect that we find (0.05) by the life expectancy bound in the Meghir et al. (2018) estimates. This effect is then divided by the decrease in our per-person life expectancy estimate from Section 6.

literature that finds associations between child neuropsychological issues, such as attention deficit hyperactivity disorder and autism spectrum disorder, and air pollution in pre- and post-natal periods (e.g., Suades-González et al., 2015) as well as recent economic studies linking sustained *later* life pollution to dementia (Bishop et al., 2018). However, to the best of our knowledge, this is the first paper to causally link early-life air pollution exposure to such difficulties in later life.

# 8 Robustness

We conduct a variety of robustness tests to ensure that our estimates do not qualitatively change when we (1) evaluate the direct effects of the fires themselves; (2) vary our specification or choose alternative treatment definitions; (3) consider the effects of mortality selection; (4) adjust for migration-related measurement error; or (5) employ alternate estimation methods. We discuss each of these exercises below.

## 8.1 Controlling for Proximity to Fires

This paper has thus far attributed the later-life harms of wildfire exposure to the substantial increases in air pollution that stem from these fires. However, wildfires sometimes cause physical damage to populated lands and may also necessitate evacuation or other responses that could result in childhood trauma. Accordingly, it is worth evaluating the degree to which our treatment may capture effects of wildfire-induced trauma, rather than harms purely from early-life air pollution exposure. To do so, we re-estimate our leads and lags specifications and include controls for fires that are within 10 miles of an individual's city of birth—specifically, individuals are assigned direct fire exposure if *any* part of the fire is within 10 miles of the city centroid. As shown by Figure 6, controlling for proximity to fires does not materially affect our estimates, suggesting that wildfire smoke—and not trauma or economic damage from the direct effects of the fire—are driving results.[44]

## 8.2 Changes to Specification

In this subsection, we include a variety of controls and vary the configuration of fixed effects in our baseline specification. Before detailing the results of these modifications, it is worth briefly discussing primary threats to internal validity and how these threats are either addressed by our main

[44] Given that loss and displacement from fires are serious events that have been shown to have lasting consequences (Schwank, 2021), it is worth considering why we *don't* see effects when controlling for these estimates. One simple explanation is that, since fires tend to occur away from population centers, very few fires are sufficiently near to people to cause lasting harm. Anecdotal evidence from historical newspaper records, which were reviewed to supplement missing information on wildfire dates, is consistent with this assertion. We noted only sparing reports of large evacuations or towns that were lost to wildfires, and even these locations tended to be sparsely populated resort or logging towns. However, it is worth noting that in *recent* years there have been several notable incidents where fires directly affected population centers. This is attributable, in part, to the rapid growth of the number of individuals living in the wildland urban interface (Radeloff et al., 2018), which both increases the exposure of populated places to naturally occurring wildfires—such as those caused by lightning strikes—*and* increases the risk that population-adjacent wildfires will be caused by human activity.

specification or can be diagnosed via our leads-and-lags coefficients. The two primary threats to internal validity would be if: (1) wildfire smoke was spuriously correlated with other characteristics and trends in a given area; or (2) wildfire smoke was co-determined with other weather-related factors—such as temperature, precipitation, or wind—that affect long-run outcomes. Our main specification addresses the first potential threat to identification by including detailed city of birth fixed effects (that finely control for time-invariant factors), along with year-of-birth $\times$ region-of-birth effects (which control for cohort-varying factors that are specific to Northern and Southern California, respectively). However, in the event that these controls do not adequately adjust for spuriously correlated time-invariant or time-varying factors, our lead coefficients ($\delta_{-5}, \delta_{-4}, \ldots, \delta_{-1}$)—which capture the marginal effect of an additional month of smoke exposure *prior* to conception—provide useful tests. If these placebo coefficients display pre-trends or are consistently and significantly different than zero, then that would indicate potential issues with our analysis.[45] However, we neither observe level shifts or trends in the placebo coefficients displayed in Sections 6 and 7 (or their related appendix figures), supporting the idea that time-invariant and time-varying factors are not confounding our analysis.

In addition to the credibility added by our placebo tests, we also further address potential threats to internal validity by adding various controls and modifying our fixed effect specifications. The results of these tests are shown in Figure 7, which displays summary estimates for our baseline specification (on the top row) along with summary estimates of alternative specifications (on other rows and described by the in-figure text). We briefly discuss each of these robustness specifications below.

**Additional Controls**

Our first specification includes controls for climate variables, which have been shown to be both a strong determinant of wildfires and also human health in the short- and long-run (e.g., Barreca et al., 2016; Isen et al., 2017a). Specifically, we include controls for the number of days above 32 degrees Celsius during the *in utero* period through the first six months of life—i.e., the critical period in which temperature strongly affects long-run earnings, as shown by Isen et al. (2017a). We also control for (a) the mean of monthly precipitation and (b) the fraction of days in cardinal wind direction bins interacted with an individual's county of birth. Both sets of these controls

[45] To provide specific examples, if there are time-varying trends (i.e., if areas with high propensity for smoke exposure are trending in a different manner than low-propensity areas), we would expect to see trends in our placebo coefficients prior to the conception date. Additionally, the lead coefficients will also detect if there is a compositional response in *anticipation* of a wildfire smoke shock. Consider if, for instance, wealthier or more educated parents—who are expected to have wealthier and more educated children—move away from an area just before their child is born. In this case, we would expect to see a preconception trend since the composition of the children born in a given area is changing over time. However, we see no trends that would suggest this is occurring. (Regarding the specific threat of anticipatory migration, we also do not find any evidence that wildfire smoke exposure induces migration, as indicated in the last column of Appendix Table A4. This is also consistent with findings from Borgschulte et al. (2020) who do not detect smoke-induced migration behavior in a modern context.)

are calculated over the *in utero* through age 4 period.[46] However, as displayed in the Figure 7, controlling for these outcomes does not qualitatively affect our estimates.

Next, in order to better control for granular time trends, we follow a standard practice in the early-life determinants literature by interacting baseline county-level characteristics with time trends (recent examples include Bailey et al., 2020, 2021; Hoynes et al., 2016). Specifically, we interact 1930 county-level data with a decade-varying-linear spline in time.[47] We find that inclusion of these interacted splines does not meaningfully change our estimates.

Last, we control for the possibility that more-and-less smoke prone areas are trending differently by dividing overall smoke propensity (defined over the entire period) into quartiles and interacting these quartiles with birth year fixed effects. Inclusion of these additional fixed effect controls do not significantly affect our estimates.

**Different Fixed Effect Specifications**

We next consider different fixed effect specifications that increase the granularity of controls along time and spatial dimensions. First, we replace our year-of-birth by region-of-birth fixed effects—where region is defined as Southern California or Northern California—with more detailed year-*month*-of-birth by region-of-birth fixed effects. If there are short-lived shocks that are contemporaneous with wildfires—such as regional temperature spikes or other extreme weather—they will be captured by these controls. Next, we instead increased the granularity of our region-of-birth fixed effects to be based on State Economic Areas (precursors to modern-day commuting zones), rather than a South-North distinction.[48] These controls will capture relatively local trends that may not be common to the larger region. As displayed in Figure 7, modifications to these fixed effects yield highly similar estimates to our baseline specification.

**Different Treatment Definition**

Our treatment definition implicitly assumes that each additional month of exposure during a given age range has the same marginal effect—i.e., that the dose-response function is linear in the number of months with *any* smoke exposure from the *in utero* period through age 4. To relax this assumption, we redefine our treatment variable to be an indicator variable equal to one if an individual is exposed to an above-median amount of smoke and zero otherwise. Specifically, for an in-sample

[46] Precipitation could affect children primarily via agricultural income shocks. Additionally, we control for county-by-wind direction bins because it is possible that wildfire smoke could blow in exclusively from an otherwise "good" or "bad" direction, i.e., a direction that has little or lots of pollution from non-wildfire sources (Deryugina et al., 2019). If this is the case, then we would systemically under (over) state the effects of wildfire smoke on long-run outcomes.

[47] These characteristics include: (1) the fraction of children in school; (2) the percent immigrant population; (3) the relative share of manufacturing and agricultural employment; (4) the average age at first marriage; and (5) the average family income. All characteristics were obtained from the 1930 Census 100% Sample (Ruggles et al., 2021), except for family income, which was obtained from the 1934 IRS Statements of Income. (No income-related questions were asked during the 1930 Census.)

[48] See Appendix Figure A11 for illustration of State Economic Areas.

median smoke measure $\bar{m}$, we estimate:

$$Y_{cba} = \mu_c + \lambda_{y(b),r(c)} + \alpha_{m(b)} + \psi_{y(b),a} + \tau \cdot \mathbf{I}\big(Smoke_c^{\text{IU-Age 4}} > \bar{m}\big) + \varepsilon_{cba}, \tag{4}$$

which is the same specification as Equation 3 with a transformed treatment variable. To facilitate comparison with our estimated treatment, we also estimate a separate equation with $Smoke_c^{\text{IU-Age 4}}$ as the *dependent* variable to obtain the average number of additional smoke months for those with above-average exposure:

$$Smoke_c^{\text{IU-Age 4}} = \mu_c + \lambda_{y(b),r(c)} + \alpha_{m(b)} + \psi_{y(b),a} + \kappa \cdot \mathbf{I}\big(Smoke_c^{\text{IU-Age 4}} > \bar{m}\big) + \varepsilon_{cba}. \tag{5}$$

We then scale the effects from Equation 4 ($\hat{\tau}$) by those from Equation 5 ($\hat{\kappa}$) to generate a comparable estimate to our main results. The scaled estimates from this exercise, presented in the final row of Figure 7, are qualitatively very similar to the effects estimated from our baseline, although they are less precise, consistent with the decreased variation in the treatment variable.

In addition to this exercise, we also provide evidence supporting the linearity of the dose-response function using residualized scatter plots. These plots, which are displayed in Appendix Figure A19, generally support the choice of a linear functional form for our analysis. See accompanying notes to these figures for further information on how plots were constructed and additional discussion of interpretation.

### 8.3 Mortality Selection

Because Section 6 demonstrates mortality effects from early childhood wildfire smoke, a potential concern, particularly for the outcomes discussed in Section 7, is that mortality selection could lead to bias in observed long-run outcomes. Put differently, it is possible that more highly exposed children will not be present in later-life surveys because they died before the survey date—either as infants or in adulthood—creating potential compositional deviations from the implied counterfactual at later ages.

To evaluate the degree to which this selection may affect estimates, we perform a bounding exercise in the same spirit as Lee (2009). This exercise is explained in detail in Appendix Section B.3, but we discuss the general method here. First, we estimate the death rate at a given age $a$ using the same methods employed in Section 6. Then, using these estimates, we predict the number of deaths for each birth cohort (defined by city and year-month of birth) by age $a$.[49] We then assign extreme values—equal to either the within-sample 1st percentile of the outcome distribution (for a lower bound) or the 99th percentile of the outcome distribution (for an upper bound)—to impute the outcome of the deceased individuals as of the survey date. We then create a new outcome variable by averaging these values into the values of other individuals observed in that birth cohort.[50] The results of this exercise are displayed in Appendix Figure A21. As demonstrated

[49] These estimates also include infant mortality, discussed in Appendix Section B.1.

[50] To give a simplified example, suppose we observe 1,000 individuals of a given cohort at age 55 with an earnings aver-

by the figure, our estimates are not sensitive to mortality selection.

### 8.4 Measurement Error

Because our estimates use place of birth to assign treatment, our treatment will be mis-measured to some degree for any individuals who migrate away from where they were born within the first four to five years of life. This is of particular concern when interpreting our age-specific estimates, as estimates at older ages will mechanically have higher rates of migration away from their birthplace and thus have higher degrees of measurement error. To assess the potential impact of measurement error on our main result, we utilize the 1940 Census 100% Sample to determine where children live at age five and how it compares to their place of birth, which is inferred using their 5-year migration status.[51] With these data, we restrict our sample to children age five who were born in California and utilize information on their counties of birth and residence to calculate measures of "true" exposure and exposure based solely on place of birth (i.e., assigned in the same way as our main analyses). Then, using these assigned treatments, we calculate how much our estimates have potentially been attenuated due to measurement error, under the assumption that error in assignment is classical. A more detailed description of this approach, as well as support for the assumption of classical measurement error, are discussed in detail in Appendix Section B.4.

The results of this exercise, where we adjust our estimates for the calculated attenuation factors, are displayed in Appendix Figure A22. Within the figure, we present the attenuation-adjusted leads-and-lags coefficients for our main outcomes from Sections 6 and 7 alongside the original estimates. As demonstrated by the figure, the measurement-error-adjusted estimates are nearly indistinguishable from our baseline estimates, likely due in part to the relatively low rates of migration among small children (only 14% of children had migrated by age 5 within the sample).

### 8.5 Issues with Two-Way Fixed Effect Estimates and Alternative Estimation

This paper utilizes two-way fixed effects ("TWFE") controls in our estimation of long-term outcomes. However, as discussed in a fast-growing literature (e.g., Callaway and Sant'Anna, 2021; de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021a; Sun and Abraham, 2021),

age of \$44,000. Additionally, assume that, based on their childhood wildfire exposure, an additional two individuals died before age 55 who would have survived had they not been exposed to wildfire smoke. We then bound the effects of these two missing individuals by assuming that they were at the 1st percentile of the earnings distribution (\$0) or at the 99th percentile of the earnings distribution (\$286,000). Then the new cohort averages are as follows:

$$\textbf{Lower}: \frac{(1,000 \times \$44,000) + (2 * \$0)}{(1000+2)} = \$43,920 \qquad \textbf{Upper}: \frac{(1,000 \times \$44,000) + (2 * \$286,000)}{(1000+2)} = \$44,483$$

These upper and lower bound averages are then used as outcome variables in our regressions.

[51] The 1940 Census is best-suited for this analysis because it asked 5-year migration questions of all individuals in the 100% sample. This contrasts with later Decennial Censuses where migration was asked only of a smaller sub-sample and/or was asked only about 1-year migration. In this context, the larger sample size is particularly useful as we are looking at relatively small sub-sample of individuals, i.e., those who were five years old in 1940 and were also born in California.

TWFE models may be difficult to interpret due to the unintuitive way in which they weight different observations across time. This can be particularly problematic if there is staggered timing, time-varying treatment effects, and/or dynamic treatment effects. Recently, several papers—such as Cengiz et al. (2019) and the previously referenced studies—have proposed methods for handling these issues in commonly used difference-in-difference and event-study frameworks where a policy is enacted and remains in effect.

However, the existing literature has not settled on the best way to deal with these weighting issues for continuous treatments that turn on and off, which is the case for our research design.[52] To address potential weighting issues, we estimate an alternate regression specification that is separately estimated *for each year-of-birth cohort* and we then combine these estimates by simple average to ensure intuitive weighting. Specifically, we estimate:

$$Y_{cba} = \lambda_{y(b),r(c)} + \alpha_{m(b)} + \psi_{y(b),a} + \delta \cdot Smoke_{cb}^{\text{IU-Age4}} + \overline{Y}_c + f(X_{cb}) + \varepsilon_{cba},$$

where the above equation is very similar to our summary specification (Equation 3), but differs in key ways to allow separate estimation for each year-of-birth cohort. Because we are estimating each regression using only within year-of-birth cohort (cross-sectional) variation, we are unable to utilize our city-of-birth fixed effects ($\mu_c$) that are present in Equation 3. To increase precision and control for potential differences across places that may be correlated with both wildfire smoke and long-run outcomes, we instead include long-run city-of-birth averages of the outcome variable ($\overline{Y_c}$) as well as indicators for quartiles of smoke propensity, county-level demographic controls, and climate controls discussed earlier in this section. The results of this exercise are presented in Appendix Figure A24. Within the figure, the estimates from our alternative estimation are qualitatively similar, though less precise—although our results pertaining to longevity, economic achievement, and educational attainment all exceed traditional levels of statistical significance.[53]

# 9 Supplemental Results

We have established that early-childhood wildfire exposure results in adverse long-run impacts to longevity, economic achievement, educational attainment, and disability in later life. In this section, we provide further evidence on (1) how effects evolve as individuals age, (2) what extent smoke exposure reshapes the distribution of outcomes, (3) potential additional mechanisms for our mortality results, and (4) other heterogeneity and outcomes that may be affected by childhood air pollution exposure.

[52] Very recent work by Callaway et al. (2021) provide alternative estimation strategies for difference-in-differences designs with continuous treatment. Future versions of this paper will consider the applicability of these methods to our setting.

[53] We obtain standard errors through 100 cluster bootstrap iterations.

### 9.1 Effects on Earnings and Disability Grow More Severe with Age

As discussed earlier in this paper, a key advantage of our setting is that we are able to observe the effects of childhood air pollution shocks for adults in later life, which enables us to better understand how early-life harms evolve throughout the life cycle. It is not *ex-ante* clear how treatment effects evolve as individuals age: the impact of childhood shocks may "fade" when transitioning from early adulthood to mid-life if, for instance, they simply accelerate the manifestation of an underlying condition that would have otherwise impacted an individual a few years later. On the other hand, it is possible that the impacts of early-life harms, mitigated by the higher "health stock" of younger individuals, *increase* proportionally as individuals age.

To assess which of these scenarios holds empirically, we evaluate the two "headline" outcomes with the potential to vary as individuals age: earnings and cognitive difficulty. Specifically, we estimate the following equation:

$$Y_{cba} = \mu_c + \lambda_{y(b),r(c)} + \alpha_{m(b)} + \psi_{y(b),a} + \sum_{j \in \mathcal{J}} \delta_j \cdot (AgeBin_j \times Smoke_c^{\text{IU-Age 4}}) + \varepsilon_{cba}, \qquad (6)$$

where $\mathcal{J}$ is the set of age bins in 5-year increments. The results of this equation are displayed in Figure 8. Within the figure, each point represents a separate estimate of $\delta_j$ with associated 95% confidence intervals. For further context, the effect as the percent of the age-specific mean is also presented, along with a line of best fit for the coefficients.

While the individual coefficients from this exercise are somewhat imprecise, taken together they tell a clear story: the adverse impact of childhood air pollution exposure on earnings and cognitive difficulty is growing with age, both in levels and proportionally. The earnings effect (as a percent of the age-specific mean) for the youngest age bin is less than one third the effect for the oldest age bin; the impact of cognitive difficulty is less than one half.[54] As discussed in Section 7.1, we utilize these life-cycle earnings dynamics to arrive at a discounted earnings loss of $4,710 in response to a $p_{25} \rightarrow p_{75}$ increase in wildfire smoke exposure. To underscore the importance of incorporating these age-specific effects into our earnings loss computation, we calculate the discounted loss of earnings using an alternative method. Due to data constrains, the existing literature has assumed that the reductions in early-adulthood earnings persist proportionally throughout the life cycle (i.e., that a 1% reduction in age 30 earnings will translate to a 1% reduction in earnings at ages 35, 45, 55, etc.). We follow this method by using the proportional reduction in earnings that we observe at our earliest age bin (aged 40 and under; $p_{25} \rightarrow p_{75}$ = 1.0%) and assume that this effect is constant as people age. Such an exercise would result in a discounted cost of air pollution that is 44% lower than our estimate—a substantial understatement. Indeed, these findings also suggest that there could also be considerable value in investigating longer-run impacts of early-life

[54] Appendix Figure A25 presents qualitatively similar results for labor force participation, economic index, and disability index outcomes. Notably, our impact on labor force participation is almost completely driven by the effects in the oldest age bin: a $p_{25} \rightarrow p_{75}$ = 1.07 percentage points. When we assess how much of the later-life earnings could be attributable to labor force participation (performing a back-of-the-envelope in the same manner as was discussed in Section 7.1), we find that reduced LFP "explains" 43.7% of the late-life earnings effect.

shocks from sources aside from air pollution.

### 9.2 Childhood Smoke Exposure Affects the Entire Distribution of Economic Achievement and Earnings

While we establish that childhood exposure to smoke reduces economic achievement and earnings *on average* (Section 7.1), it is not clear how this exposure is reshaping the earnings distribution. To assess this, we calculate the in-sample quintiles of the economic index and earnings distribution. We then estimate separate regressions where the outcome variable is equal to one if an individual is in a given quintile bin and zero otherwise.

The results of this exercise are displayed in Figure 9 where each bar represents the coefficient from a separate regression and associated 95% confidence intervals. Both Panels A and B tell a similar story: early-life exposure to wildfire pollution meaningfully increases the probability that individuals are in the worst quintile of the economic achievement and earnings distribution ($p_{25} \rightarrow p_{75}$ = 0.6 to 0.7 percentage points) and that impact is almost completely offset by reductions in the best quintile ($p_{25} \rightarrow p_{75}$ = 0.5 to 0.6pp). These results are consistent with a left shift of the distribution of outcomes. Our findings also draw an interesting contrast with those of Isen et al. (2017b) who perform a very similar exercise to ascertain the distributional impact of the Clean Air Act Amendments of 1970 ("CAAA"). While they find substantial improvements to the bottom of the income distribution (which is analogous to our findings), their effect is largely displaced by increases in mass among the middle quantiles.

### 9.3 Early-Life Exposure to Wildfire Smoke Induces Individuals to Sort into Higher-Mortality Occupations and Neighborhoods

Our analyses in Section 7 of this paper suggest that a substantial portion of the mortality effects that we find could be attributable to reductions in economic achievement and educational attainment. One potential way in which this could occur is if individuals are induced into occupations or neighborhoods that increase their risk of death via environmental exposures, occupational safety hazards, increased levels of stress, higher rates of violent crime, or other factors. To further explore this possibility, we examine the additional mortality risk that is associated with individuals' choice of occupation and neighborhood. Specifically, we utilize the full Numident file (i.e., without restricting to Californian births) and link it to both the full American Community Surveys and 2000/2010 Decennial Censuses. Using these data, we then restrict to the same age ranges that are covered within this paper and calculate 10-year mortality rates for each occupation and Census tract.[55] We then assign occupation-specific mortality rates to individuals based on their observed occupation in the ACS and assign tract-specific mortality rates to individuals based on their residence in the Decennial Census.

[55] Occupation-specific mortality rates are also within sex and 5-year age bin, while tract-specific mortality rates are within sex.

Having constructed these ten-year mortality measures, we utilize them as outcome variables to assess whether childhood smoke exposure causes individuals to sort into higher-mortality occupations and neighborhoods. The results of these regressions, displayed in Figure 10, clearly demonstrate that more exposed individuals sort into occupations and neighborhoods with greater mortality risk. While this result does not conclusively show that longevity is increasing due to occupation and neighborhood choice—after all, lower socioeconomic status exposes individuals to a number of increased mortality risks that could also be correlated with these choices—it does provide suggestive evidence as to the channels that could contribute to premature mortality.[56]

### 9.4 Wildfire Shocks Are Not Transmitted to Children Via Local Economic Channels and Long-Run Effects Do Not Meaningfully Vary Across Person and Place Characteristics

We perform two other supplemental analysis, both of which are discussed further in Appendix Section B.5 but are briefly summarized here. First, we perform an analysis of contemporaneous economic outcomes (e.g., employment and family income) and net migration to ascertain if wildfire smoke shocks could be transmitted to children via local economic channels (see Appendix Table A4). We do not find any statistically significant evidence of reductions in economic activity. Further, the reductions that we do find are also economically small and cannot explain a substantial portion of the longevity results that we find.

Second, we consider heterogeneity across individual-level and place-level characteristics, including sex, race/ethnicity, urban status, county-level income, hospital access, and county-level economic dependence on lumber (i.e., places that could be more sensitive to loss of timber due to wildfire). The results of this analysis, shown in Appendix Figure A26, display little-to-no heterogeneity. Out of our twenty-four regressions (six groups $\times$ four outcomes), there are only three that have statistically different interaction effects.

### 9.5 The Effect of Childhood Wildfire Exposure on Family Formation, Institutionalization, and Other Outcomes

In addition to the wide set of outcomes explored in previous sections, we briefly discuss certain outcomes here that do not fall into one of our three indices but are of interest. These outcomes, along with their scaled coefficients (as a percent of the mean), are presented in Appendix Figure A27. As demonstrated by the figure, the effects that we estimate for family formation outcomes (whether a person has ever been married and the number of children they have living with them at the time of survey) and migration are either statistically insignificant, economically insignificant, or both. Likewise, we estimate the effect of early-childhood wildfire smoke on two placebo outcomes—sex and race/ethnicity—which should not be affected by air pollution. As expected, we do not detect any impact on these outcomes.

[56] Appendix Figure A18, Panel B, also considers the 10-year mortality rate of an individual's industry and finds qualitatively similar findings to those discussed here.

We do, however, see increases in the number of individuals who are held in institutional group quarters—i.e., correctional or nursing facilities—as of the 2010 Decennial Census.[57] Specifically, we find that moving from the 25$^{th}$ to 75$^{th}$ percentile of exposure increases institutionalization by 0.025 percentage points, a 2.1% increase over the mean. This finding is consistent with the literature finding early-life-exposure impacts of lead (e.g., Reyes, 2007; Aizer and Currie, 2019; Grönqvist et al., 2020) and other particulate matter (Voorheis, 2017) on incarceration and behavioral issues. Further, it underscores the ability of wildfire pollution exposure to cause the sort of degradation in health and non-cognitive skills that ultimately leads to institutionalization.

# 10 Discussion and Conclusion

## 10.1 The Cost and Fiscal Burden of Childhood Wildfire Smoke Exposure

We end our analysis by calculating the cost of wildfires to children in our sample by multiplying our estimates by the average exposure of each child, which gives the per-person benefit of "shutting off" wildfire pollution. Additionally, we also quantify the fiscal burden by estimating the amount of lost tax revenue as well as increases in social safety net costs and incarceration costs as a result of wildfires.

The results of this exercise are presented in Table 1. The left half of the table (Panel A) details our estimates of the individual cost of wildfire pollution in two areas: earnings and life years lost in adulthood. Using the same discounting method discussed in Sections 7.1 and 9.1, we calculate the discounted losses for the average amount of *in utero* through age 4 wildfire smoke exposure within our sample (10.5 months) and find a reduction of approximately \$9,900 per person. Added to this amount is the estimated cost of lost life years in adulthood, calculated in Section 6. Specifically, we estimate the number of lost life years $\left(\frac{9.2 \text{ life years lost}}{1{,}000 \text{ persons}} \times 10.5 \text{ smoke months}\right)$ by the estimated value of a statistical life year of \$130,000 to obtain a per-person mortality cost of roughly \$12,600.[58] Taken together, these estimates imply per-person losses of \$22,450 due to wildfire smoke exposure. To provide further perspective on the magnitude of these losses, we multiply them by 500,000 persons, which is roughly the size of a single California birth cohort. The exercise yields losses of \$11.22 billion dollars for every single "cohort" subjected to this average level of exposure. These estimates, while large, conservatively omit several components. First, because of this paper's focus on *later* life outcomes, we omit any earnings losses before age 35 and any losses due to infant mortality. While we do not find statistically significant early-adult earnings effects, incorporation of infant

[57] This variable is not available in the Restricted 2000 Decennial Census. Note that correctional and nursing facilities are grouped together because the Census Bureau did not permit RDC researchers to disclose results using sub-categories of the institutional group quarters variable during the time period in which these results were disclosed. However, as approximately 85.2% of institutionalized persons in our sample age range were in correctional facilities, it is reasonable to interpret these effects as primarily pertaining to incarceration.

[58] The value of a statistical life year ("VLSY") was obtained from Cutler and Richardson (1999) and inflated to 2010 dollars. Alternatively, we could utilize the methods employed by Carleton et al. (2020) and Bailey et al. (2020). To determine the VLSY, these studies use the *age-invariant* value of statistical life estimates from the Environmental Protection Agency and divide them by median life expectancy in the United States to arrive an approximate value of \$200,000 (in 2010\$). If we instead utilized these estimates, the life years lost would be valued at \$19,320.

mortality costs would increase the estimate by $1.6 or $2.5 billion dollars. Additionally, we do not attempt to quantify the reduced quality of life from increased disability and the lost non-pecuniary benefits due to lower educational attainment (Oreopoulos and Salvanes, 2011).

In addition to the direct, person-level costs of wildfire smoke exposure, there are also increased fiscal burdens on governments via lower tax revenue, increased use of the social safety net, and increased incarceration. These fiscal burdens are displayed in Panel B of Table 1. The first row, which is concerned with lower tax revenue, takes the decreased earnings (from Panel A) and multiplies them by the average federal marginal tax rate from 2000-2019 of 13.2% (Tax Policy Center, 2021) to arrive at the lost tax revenue.[59] The cost impact of increased social safety net use is taken from Section 7.1 and scaled by the average exposure of 10.5 months to arrive at a discounted burden of $482 per person. Finally, we utilize the increases in incarceration from the previous section to calculate the burden of housing inmates. Using costs from Vera Institute of Justice (2015, 2012), Bureau of Justice Statistics (2016), and Sawyer and Wagner (2020), we calculate the per-year cost of incarceration and discount back to birth.[60] This exercise yields an average discounted cost of $99 per person. When combined, we calculate that these factors impose an additional cost of $1,887 per person, or $0.94 billion per 500,000 individuals.

These impacts were estimated using historical birth cohorts from the 1930s through 1960s. However, it is worth considering whether these effects are expected for children born in most recent years. On one hand, there are substantial advances in medical technology available to modern children that were not available during our sample period, and there is an increased ability to avoid air pollution due to indoor air filtration. However, the degree to which these technologies are able to offset the lasting effects of early-life shocks is unknown: insofar as impacts to childhood health are not salient—which might be the case for mild cognitive impairment or exposures that increase long-run cancer risk, for example—it is unclear whether modern technology provides substantial protection from long-run harms. Furthermore, it is possible that any improvements in our ability to lessen or avoid the impacts of childhood wildfire smoke may be more than offset by the increasing severity of wildfire seasons (Figure 1), which subject children to more frequent and more severe smoke exposure.[61] Because the net effect of these countervailing factors are unclear, we believe that more research on the early childhood impacts of wildfire smoke in a modern context is needed. A potential advantage in studying wildfire pollution is that, unlike industrial pollution

[59] This exercise abstracts away from losses in state and local tax revenue, but these could also be substantial as lower earnings likely also translate into reduced sales taxes via decreased consumption.

[60] Specifically, we gather incarceration costs for local jails (Vera Institute of Justice, 2015), state prisons (Vera Institute of Justice, 2012), and federal prisons (Bureau of Justice Statistics, 2016), converting them to 2010 dollars. We then weight them by the relative number of individuals in each level of incarceration (Sawyer and Wagner, 2020) to obtain the cost per incarceration year. We further adjust this estimate to reflect that we use the "institutionalized group quarters" variable from the Census, which also includes individuals in nursing facilities. (As discussed in the previous section, we cannot separately disclose incarceration results due to disclosure restrictions.) Accordingly, we down-weight our cost estimate by the fraction of institutionalized 35-59 year-olds that were in prison (85.2%). We then use this down-weighted incarceration cost measure and discount back to birth, using the conservative assumption that individuals are equally likely to be incarcerated at any given year in this age range.

[61] Future iterations of this paper will quantify the degree to which present-day children are more exposed to wildfire smoke than their historical counterparts in our sample period.

sources such as coal power plants, the emissions profile has likely not substantially changed over time. Accordingly, any studies that determine the medium-run effects of early-life smoke exposure within a more recent context (using outcomes such as education or early-adult earnings) will be able to compare findings to estimates in this paper. Such comparisons may improve our ability to understand how the impacts of these childhood shocks have evolved over time.

### 10.2 Conclusion

In this paper, we contribute to the understanding of the long-term effects of childhood air pollution exposure by leveraging variation in an increasingly important source: smoke from wildfires. Specifically, we examine longevity, a previously unstudied outcome in the literature. We also provide a deeper understanding of the impact of early-life exposure across the life cycle by evaluating economic, education, and disability outcomes at later ages. Specifically, we find that moving from the $25^{th}$ to the $75^{th}$ percentile of smoke exposure results in 46 life years lost per 1,000 persons and reduces annual earnings by $890. We further find that this level of exposure reduces the probability of completing high school (by 0.40 percentage points) and increases the probability of cognitive difficulty (by 0.30 percentage points). Further investigation of these later-life harms demonstrates that they grow proportionally worse as individuals age and that failure to incorporate these later-life dynamics can potentially cause large understatements in the discounted damages from early-life shocks.

Because the findings in this paper demonstrate that childhood exposure to wildfire pollution has substantial long-term costs, there are potentially larger benefits than previously understood from policies which abate, ameliorate, or defend against the harms from wildfire smoke. We conclude by briefly discussing these policies. First, because worsening wildfires—and the amounts of related air pollution—are exacerbated by the warming climate (Vose et al., 2012; IPCC, 2021; Zhuang et al., 2021), any policy that slows or reverses the trend of higher temperatures will have an associated benefit of improved health from decreased wildfire smoke. However, absent a reversal of warming temperatures, it is possible that improved forest management practices could reduce the air pollution from wildfires. A substantial amount of resources in fire prevention and containment are devoted to protecting homes and other structures (Plantinga et al., 2020); however, the costs of wildfire pollution suggest that minimizing the biomass burned is important even when fires occur in remote areas.

One potential policy that could abate the occurrence of large wildfires is an increased use of prescribed burns, a technique that has potentially been underutilized in California and other western states (Boxall, 2021; Sommer, 2021). Traditionally, local and state authorities have had substantial responsibility for enacting prescribed burns. However, because wildfire smoke can create meaningful interstate health externalities, it is possible that Federal regulators should take a larger role in encouraging forest management practices that prevent large fires. Schultz et al. (2019) suggests that adequate funding and incentives are key barriers to sufficient use of prescribed burns, two

areas that could benefit from an increased role by the Federal government.[62]

While abatement of wildfires could be an important policy lever, it is not the only option for reducing harms of childhood pollution. Because the impacts of wildfire smoke on indoor air quality are much lower (Burke et al., 2021), policies that encourage indoor behavior and improve indoor air filtration in homes, schools, and workplaces may be beneficial. Although more research on the effectiveness of filtration against wildfire smoke is needed, existing studies suggest that improving general air filtration, especially in schools, is very cost-effective (e.g., Gilraine, 2020; Stafford, 2015).[63] Finally, we note that it may be possible to mitigate injury from childhood pollution shocks even after children are exposed (Billings and Schnepel, 2018). Absent interventions to reduce or avoid wildfire smoke through other methods, understanding ways to reduce the harms of air pollution through medical treatment, social programs, or other means is an important area for future research.

[62] Of course, pollution from controlled burns can be harmful as well. However, controlled burns consume much smaller amounts of biomass and can be planned on days when pollution would be blown away from densely-populated areas.

[63] Less is known about the impact of air filtration in homes. Burke et al. (2021) note that there is substantial dispersion in the degree of smoke infiltration among different homes, but that household characteristics, such as income, do not explain much of the variation.

# References


Aguilera, R., T. Corringham, A. Gershunov, and T. Benmarhnia (2021a). Wildfire Smoke Impacts Respiratory Health More Than Fine Particles from Other Sources: Observational Evidence from Southern California. *Nature Communications 12*(1), 1493.

Aguilera, R., T. Corringham, A. Gershunov, S. Leibel, and T. Benmarhnia (2021b). Fine Particles in Wildfire Smoke and Pediatric Respiratory Health in California. *Pediatrics 147*(4).

Aizer, A. and J. Currie (2019). Lead and Juvenile Delinquency: New Evidence from Linked Birth, School, and Juvenile Detention Records. *Review of Economics and Statistics 101*(4), 575–587.

Aizer, A., S. Eli, J. Ferrie, and A. Lleras-Muney (2016). The Long-Run Impact of Cash Transfers to Poor Families. *American Economic Review 106*(4), 935–71.

Almond, D., J. Currie, and V. Duque (2018). Childhood Circumstances and Adult Outcomes: Act II. *Journal of Economic Literature 56*(4), 1360–1446.

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association 103*(484), 1481–1495.

Bailey, M., K. Clay, P. Fishback, M. R. Haines, S. Kantor, E. Severnini, and A. Wentz (2018). U.S. County-Level Natality and Mortality Data, 1915-2007. Inter-university Consortium for Political and Social Research [distributor].

Bailey, M. J., H. W. Hoynes, M. Rossin-Slater, and R. Walker (2020). Is the Social Safety Net a Long-Term Investment? Large-scale Evidence from the Food Stamps Program. Working Paper 26942, National Bureau of Economic Research.

Bailey, M. J., S. Sun, and B. D. Timpe (2021). Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency. Working Paper 28268, National Bureau of Economic Research.

Barreca, A., K. Clay, O. Deschenes, M. Greenstone, and J. S. Shapiro (2016). Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship over the Twentieth Century. *Journal of Political Economy 124*(1), 105–159.

Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive Linear Step-up Procedures That Control the False Discovery Rate. *Biometrika 93*(3), 491–507.

Benjamini, Y. and D. Yekutieli (2001). The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of Statistics 29*(4), 1165–1188.

Billings, S. B. and K. T. Schnepel (2018, July). Life after Lead: Effects of Early Interventions for Children Exposed to Lead. *American Economic Journal: Applied Economics 10*(3), 315–44.

Bishop, K. C., J. D. Ketcham, and N. V. Kuminoff (2018). Hazed and Confused: The Effect of Air Pollution on Dementia. Working Paper 24970, National Bureau of Economic Research.

Borgschulte, M., D. Molitor, and E. Zou (2020). Air Pollution and the Labor Market: Evidence from Wildfire Smoke. Working paper. Retrieved from http://www.davidmolitor.com/research [Date Accessed: February 22, 2021].

Boxall, B. (2021). Billions of Dollars Spent on Fighting California Wildfires, but Little on Prevention. *The Los Angeles Times*.

Buckles, K. S. and D. M. Hungerman (2013). Season of Birth and Later Outcomes: Old Questions, New Answers. *Review of Economics and Statistics 95*(3), 711–724.

Bureau of Justice Statistics (2016). Annual Determination of Average Cost of Incarceration. *Federal Register* 81 (July 19, 2016): 46957.

Burke, M., S. Heft-Neal, J. Li, A. Driscoll, P. W. Baylis, M. Stigler, J. Weill, J. Burney, J. Wen, M. Childs, and C. Gould (2021). Exposures and Behavioral Responses to Wildfire Smoke. Working Paper 29380, National Bureau of Economic Research.

California Air Resources Board (2021). Camp Fire Air Quality Data Analysis. Technical report. https://ww2.arb.ca.gov/resources/documents/camp-fire-air-quality-data-analysis (accessed August 17, 2021).

California Department of Forestry and Fire Protection (2018). Fire Perimeters - California [ds396]. https://map.dfg.ca.gov/metadata/ds0396.html [Date Accessed: October 18, 2018].

Callaway, B., A. Goodman-Bacon, and P. H. C. Sant'Anna (2021). Difference-in-Differences with a Continuous Treatment. Working paper, arXiv.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics 225*(2), 200–230. Themed Issue: Treatment Effect 1.

Card, D. (1999). The Causal Effect of Education on Earnings. In O. Ashenfelter, D. Card, and R. Layard (Eds.), *Handbook of Labor Economics*, Volume 3A, Chapter 30, pp. 1801–1863. Amsterdam: Elsevier Science & Technology.

Carleton, T. A., A. Jina, M. T. Delgado, M. Greenstone, T. Houser, S. M. Hsiang, A. Hultgren, R. E. Kopp, K. E. McCusker, I. B. Nath, J. Rising, A. Rode, H. K. Seo, A. Viaene, J. Yuan, and A. T. Zhang (2020). Valuing the Global Mortality Consequences of Climate Change Accounting for Adaptation Costs and Benefits. Working Paper 27599, National Bureau of Economic Research.

Case, A. and C. Paxson (2010). Causes and Consequences of Early-Life Health. *Demography 47*(1), S65–S85.

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019). The Effect of Minimum Wages on Low-Wage Jobs*. *The Quarterly Journal of Economics 134*(3), 1405–1454.

Chen, Y., A. Ebenstein, M. Greenstone, and H. Li (2013). Evidence On the Impact of Sustained Exposure to Air Pollution on Life Expectancy from China's Huai River Policy. *Proceedings of the National Academy of Sciences 110*(32), 12936–12941.

Chetty, R., M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler (2016). The Association Between Income and Life Expectancy in the United States, 2001-2014. *JAMA 315*(16), 1750–1766.

Clark, D. and H. Royer (2013). The Effect of Education on Adult Mortality and Health: Evidence from Britain. *American Economic Review 103*(6), 2087–2120.

Clay, K., J. Lewis, and E. Severnini (2016). Canary in a Coal Mine: Infant Mortality Tradeoffs Associated with Mid-20th Century Air Pollution. Working Paper 22155, National Bureau of Economic Research.

Colmer, J. and J. Voorheis (2020). The Grandkids Aren't Alright: The Intergenerational Effects of Prenatal Pollution Exposure. Technical report, Working Paper.

Colmer, J., J. Voorheis, and B. Williams (2022). Air Pollution and Economic Opportunity in the United States. Technical report, Working Paper.

Cropper, M. L., N. Z. Muller, Y. Park, and V. Perez-Zetune (2022). The Impact of the Clean Air Act on Particulate Matter in the 1970s. Working Paper 30312, National Bureau of Economic Research.

Currie, J. and D. Almond (2011). Human Capital Development Before Age Five. In *Handbook of Labor Economics*, Volume 4, pp. 1315–1486. Elsevier.

Cutler, D. M. and E. Richardson (1999). Your Money and Your Life: The Value of Health and What Affects It. In A. M. Garber (Ed.), *Frontiers in Health Policy Research*, Volume 2, pp. 99–132. MIT.

de Chaisemartin, C. and X. D'Haultfœuille (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review 110*(9), 2964–96.

Deryugina, T., G. Heutel, N. H. Miller, D. Molitor, and J. Reif (2019). The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction. *American Economic Review 109*(12), 4178–4219.

Finkelstein, A. (2007). The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare. *The Quarterly Journal of Economics 122*(1), 1–37.

Finlay, K. and K. R. Genadek (2021). Measuring All-Cause Mortality With the Census Numident File. *American Journal of Public Health 111*(S2), S141–S148. PMID: 34314212.

Flood, S., M. King, R. Rodgers, S. Ruggles, J. R. Warren, and M. Westberry (2021). IPUMS, Current Population Survey: Version 9.0 [dataset]. Minneapolis, MN: IPUMS, 2021. https://doi.org/10.18128/D030.V9.0.

Flynn, P. and M. M. Marcus (2021). A Watershed Moment: The Clean Water Act and Infant Health. Working Paper 29152, National Bureau of Economic Research.

Gibson, M. (2019). Regulation-Induced Pollution Substitution. *The Review of Economics and Statistics 101*(5), 827–840.

Gilraine, M. (2020). Air Filters, Pollution and Student Achievement. Working Paper 20-188, Annenberg Institute at Brown University.

Goldstein, J. R., M. Alexander, C. Breen, A. M. González, F. Menares, M. Osborne, M. Snyder, and U. Yildirim (2021). CenSoc Mortality File: Version 2.0.

Goodman-Bacon, A. (2021a). Difference-in-Differences with Variation in Treatment Timing. *Journal of Econometrics 225*(2), 254–277. Themed Issue: Treatment Effect 1.

Goodman-Bacon, A. (2021b). The Long-Run Effects of Childhood Insurance Coverage: Medicaid Implementation, Adult Health, and Labor Market Outcomes. *American Economic Review 111*(8), 2550–93.

Graff Zivin, J. and M. Neidell (2014). Temperature and the Allocation of Time: Implications for Climate Change. *Journal of Labor Economics 32*(1), 1–26.

Grainger, C. and T. Ruangmas (Nov 2018). Who Wins from Emissions Trading? Evidence from California. *Environmental and Resource Economics 71*(3), 703–727.

Grainger, C., A. Schreiber, and W. Chang (2017). Do Regulators Strategically Avoid Pollution Hotspots When Siting Monitors? Evidence from Remote Sensing of Air Pollution. Working paper.

Greenstone, M. (2002). The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures. *Journal of Political Economy 110*(6), 1175–1219.

Grönqvist, H., J. P. Nilsson, and P.-O. Robling (2020). Understanding How Low Levels of Early Lead Exposure Affect Children's Life Trajectories. *Journal of Political Economy 128*(9), 3376–3433.

Holm, S. M., M. D. Miller, and J. R. Balmes (2021). Health Effects of Wildfire Smoke in Children and Public Health Tools: a Narrative Review. *Journal of Exposure Science & Environmental Epidemiology 31*(1), 1–20.

Hoynes, H., D. W. Schanzenbach, and D. Almond (2016). Long-Run Impacts of Childhood Access to the Safety Net. *American Economic Review 106*(4), 903–34.

Hsiang, S., R. Kopp, A. Jina, J. Rising, M. Delgado, S. Mohan, D. J. Rasmussen, R. Muir-Wood, P. Wilson, M. Oppenheimer, K. Larsen, and T. Houser (2017). Estimating Economic Damage from Climate Change in the United States. *Science 356*(6345), 1362–1369.

IPCC (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Working paper. [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou (eds.)]. Cambridge University Press.

Isen, A., M. Rossin-Slater, and R. Walker (2017a). Relationship Between Season of Birth, Temperature Exposure, and Later Life Wellbeing. *Proceedings of the National Academy of Sciences 114*(51), 13447–13452.

Isen, A., M. Rossin-Slater, and W. R. Walker (2017b). Every Breath You Take — Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970. *Journal of Political Economy*.

Jacob, B., L. Lefgren, and E. Moretti (2007). The Dynamics of Criminal Behavior Evidence from Weather Shocks. *Journal of Human resources 42*(3), 489–527.

Jayachandran, S. (2009). Air Quality and Early-Life Mortality: Evidence from Indonesia's Wildfires. *Journal of Human Resources 44*(4), 916–954.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental Analysis of Neighborhood Effects. *Econometrica 75*(1), 83–119.

Landguth, E. L., Z. A. Holden, J. Graham, B. Stark, E. B. Mokhtari, E. Kaleczyc, S. Anderson, S. Urbanski, M. Jolly, E. O. Semmens, D. A. Warren, A. Swanson, E. Stone, and C. Noonan (2020). The Delayed Effect of Wildfire Season Particulate Matter on Subsequent Influenza Season in a Mountain West Region of the USA. *Environment International 139*, 105668.

Larkin, N. K., S. M. O'Neill, R. Solomon, S. Raffuse, T. Strand, D. C. Sullivan, C. Krull, M. Rorig, J. Peterson, and S. A. Ferguson (2009). The BlueSky Smoke Modeling Framework. *International Journal of Wildland Fire 18*, 906–920.

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies 76*(3), 1071–1102.

Liang, Y., D. Sengupta, M. J. Campmier, D. M. Lunderberg, J. S. Apte, and A. H. Goldstein (2021). Wildfire Smoke Impacts on Indoor Air Quality Assessed Using Crowdsourced Data in California. *Proceedings of the National Academy of Sciences 118*(36).

Logan, J. and W. Zhang (2020). Developing GIS Maps for U.S. Cities in 1930 and 1940. In I. Gregory, D. DeBats, and D. Lafreniere (Eds.), *The Routledge Companion to Spatial History*. UK: Routledge.

Manson, S., J. Schroeder, D. Van Riper, T. Kugler, and S. Ruggles (2021). IPUMS National Historical Geographic Information System: Version 16.0 [dataset]. Minneapolis, MN. http://doi.org/10.18128/D050.V16.0.

Meghir, C., M. Palme, and E. Simeonova (2018). Education and Mortality: Evidence from a Social Experiment. *American Economic Journal: Applied Economics 10*(2), 234–56.

Miller, N., D. Molitor, and E. Zou (2017). Blowing Smoke: Health Impacts of Wildfire Plume Dynamics. Working paper. Retrieved from http://www.eric-zou.com/ [Date Accessed: December 2, 2018].

Moritz, M. A., M.-A. Parisien, E. Batllori, M. A. Krawchuk, J. Van Dorn, D. J. Ganz, and K. Hayhoe (2012). Climate Change and Disruptions to Global Fire Activity. *Ecosphere 3*(6), art49.

Oreopoulos, P. and K. G. Salvanes (2011). Priceless: The Nonpecuniary Benefits of Schooling. *Journal of Economic Perspectives 25*(1), 159–84.

Peeples, L. (2020). How Air Pollution Threatens Brain Health. *Proceedings of the National Academy of Sciences 117*(25), 13856–13860.

Plantinga, A. J., R. Walsh, and M. Wibbenmeyer (2020). Priorities and Effectiveness in Wildfire Management: Evidence from Fire Spread in the Western United States. Technical Report 20-21, Resources for the Future.

Radeloff, V. C., D. P. Helmers, H. A. Kramer, M. H. Mockrin, P. M. Alexandre, A. Bar-Massada, V. Butsic, T. J. Hawbaker, S. Martinuzzi, A. D. Syphard, and S. I. Stewart (2018). Rapid Growth of the US Wildland-Urban Interface Raises Wildfire Risk. *Proceedings of the National Academy of Sciences 115*(13), 3314–3319.

Rangel, M. A. and T. Vogl (2019). Agricultural Fires and Health at Birth. *Review of Economics and Statistics*.

Reyes, J. W. (2007). Environmental Policy as Social Policy? The Impact of Childhood Lead Exposure on Crime. *The B.E. Journal of Economic Analysis & Policy 7*(1).

Reyes, J. W. (2015). Lead Exposure and Behavior: Effects on Antisocial and Risky Behavior Among Children and Adolescents. *Economic Inquiry 53*(3), 1580–1605.

Ruggles, S., S. Flood, S. Foster, R. Goeken, J. Pacas, M. Schouweiler, and M. Sobek (2021). IPUMS USA: Version 11.0 [dataset]. Minneapolis, MN: IPUMS, 2021. https://doi.org/10.18128/D010.V11.0.

Russ, T. C., M. P. C. Cherrie, C. Dibben, S. Tomlinson, S. Reis, U. Dragosits, M. Vieno, R. Beck, E. Carnell, N. K. Shortt, G. Muniz-Terrera, P. Redmond, A. M. Taylor, T. Clemens, M. van Tongeren, R. M. Agius, J. M. Starr, I. J. Deary, and J. R. Pearce (2021). Life Course Air Pollution Exposure and Cognitive Decline: Modelled Historical Air Pollution Data and the Lothian Birth Cohort 1936. *Journal of Alzheimer's disease : JAD 79*(3), 1063–1074.

Sanders, N. J. (2012). What Doesn't Kill You Makes You Weaker Prenatal Pollution Exposure and Educational Outcomes. *Journal of Human Resources 47*(3), 826–850.

Sawyer, W. and P. Wagner (2020). Mass Incarceration: The Whole Pie 2020. Technical report, Prison Policy Initiative.

Schlenker, W. and M. J. Roberts (2009). Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change. *Proceedings of the National Academy of Sciences 106*(37), 15594–15598.

Schlenker, W. and W. R. Walker (2015). Airports, Air Pollution, and Contemporaneous Health. *The Review of Economic Studies 83*(2), 768–809.

Schraufnagel, D. E. (2020). The Health Effects of Ultrafine Particles. *Experimental & Molecular Medicine 52*(3), 311–317.

Schultz, C. A., S. M. McCaffrey, and H. R. Huber-Stearns (2019). Policy Barriers and Opportunities for Prescribed Fire Application in the Western United States. *International Journal of Wildland Fire 28*(11), 874–884.

Schwank, H. (2021). Disruptive Effects of Natural Disasters: The 1906 San Francisco Fire. Working paper. https://sites.google.com/bu.edu/hannaschwank/research (accessed November 28, 2021).

Sommer, L. (2021). Why The South Is Decades Ahead Of The West In Wildfire Prevention. https://www.npr.org/2021/08/31/1029821831/to-stop-extreme-wildfires-california-is-learning-from-florida (retrieved October 30, 2021).

Stafford, T. M. (2015). Indoor Air Quality and Academic Performance. *Journal of Environmental Economics and Management 70*, 34–50.

Stein, A. F., R. R. Draxler, G. D. Rolph, B. J. B. Stunder, M. D. Cohen, and F. Ngan (2015). NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System. *Bulletin of the American Meteorological Society 96*(12), 2059 – 2077.

Suades-González, E., M. Gascon, M. Guxens, and J. Sunyer (2015). Air Pollution and Neuropsychological Development: A Review of the Latest Evidence. *Endocrinology 156*(10), 3473–3482.

Sullivan, D. (2017). The True Cost of Air Pollution: Evidence from the Housing Market. Working paper. Retrieved from http://www.danielmsullivan.com/ [Date Accessed: October 7, 2021].

Sun, L. and S. Abraham (2021). Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. *Journal of Econometrics 225*(2), 175–199. Themed Issue: Treatment Effect 1.

Tax Policy Center (2021). Historical Average Federal Tax Rates for All Households. https://www.taxpolicycenter.org/statistics/percent-tax-filers-marginal-tax-rate [accessed October 22, 2021.

U.S. Census Bureau (2012). County and City Data Book [United States] Consolidated File: County Data, 1947-1977. Inter-university Consortium for Political and Social Research [distributor]. 10.3886/ICPSR07736.v2.

Vera Institute of Justice (2012). The Price of Prisons: What Incarceration Costs Taxpayers. Technical report. Technical report, Center on Sentencing and Corrections.

Vera Institute of Justice (2015). The Price of Jails: Measuring the Taxpayer Cost of Local Incarceration. Technical report. Technical report, Center on Sentencing and Corrections.

Voorheis, J. (2017). Air Quality, Human Capital Formation and the Long-term Effects of Environmental Inequality at Birth. Technical report, CARRA Working Papers 2017-05, Center for Economic Studies, U.S. Census Bureau.

Vose, J. M., D. L. Peterson, and T. Patel-Weynand (2012). Effects of Climatic Variability and Change on Forest Ecosystems: A Comprehensive Science Synthesis for the U.S. Forest Sector. Gen. tech. rep. pnw-gtr-870. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Walker, W. R. (2013). The Transitional Costs of Sectoral Reallocation: Evidence From the Clean Air Act and the Workforce*. *The Quarterly Journal of Economics 128*(4), 1787–1835.

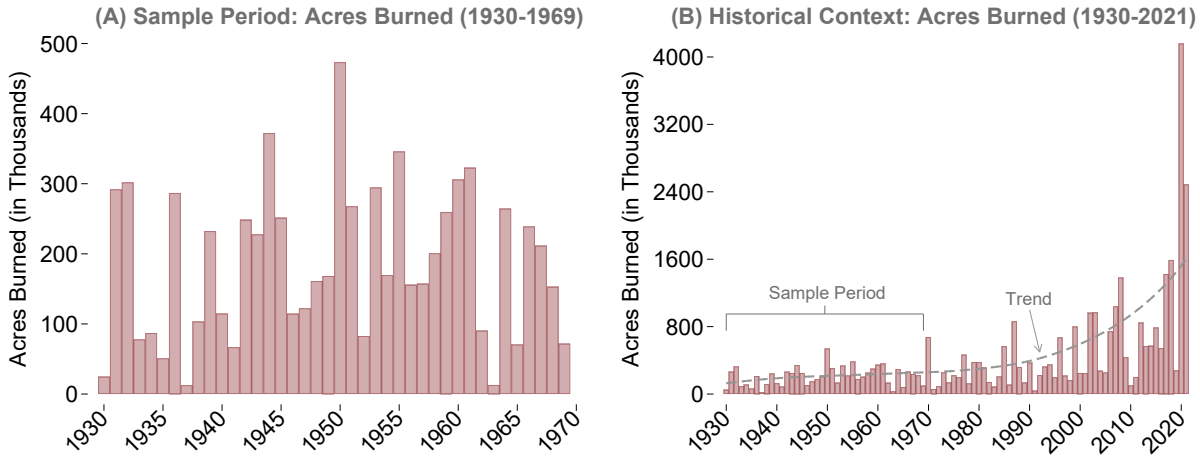
Wegesser, T. C., L. M. Franzi, F. M. Mitloehner, A. Eiguren-Fernandez, and J. A. Last (2010). Lung Antioxidant and Cytokine Responses to Coarse and Fine Particulate Matter from the Great California Wildfires of 2008. *Inhalation Toxicology 22*(7), 561–570.

Willmott, C. J. and K. Matsuura (2001). Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1950 - 1999).

Zhuang, Y., R. Fu, B. D. Santer, R. E. Dickinson, and A. Hall (2021). Quantifying Contributions of Natural Variability and Anthropogenic Forcings on Increased Fire Weather Risk over the Western United States. *Proceedings of the National Academy of Sciences 118*(45).

Zou, E. Y. (2021). Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality. *American Economic Review 111*(7), 2101–26.

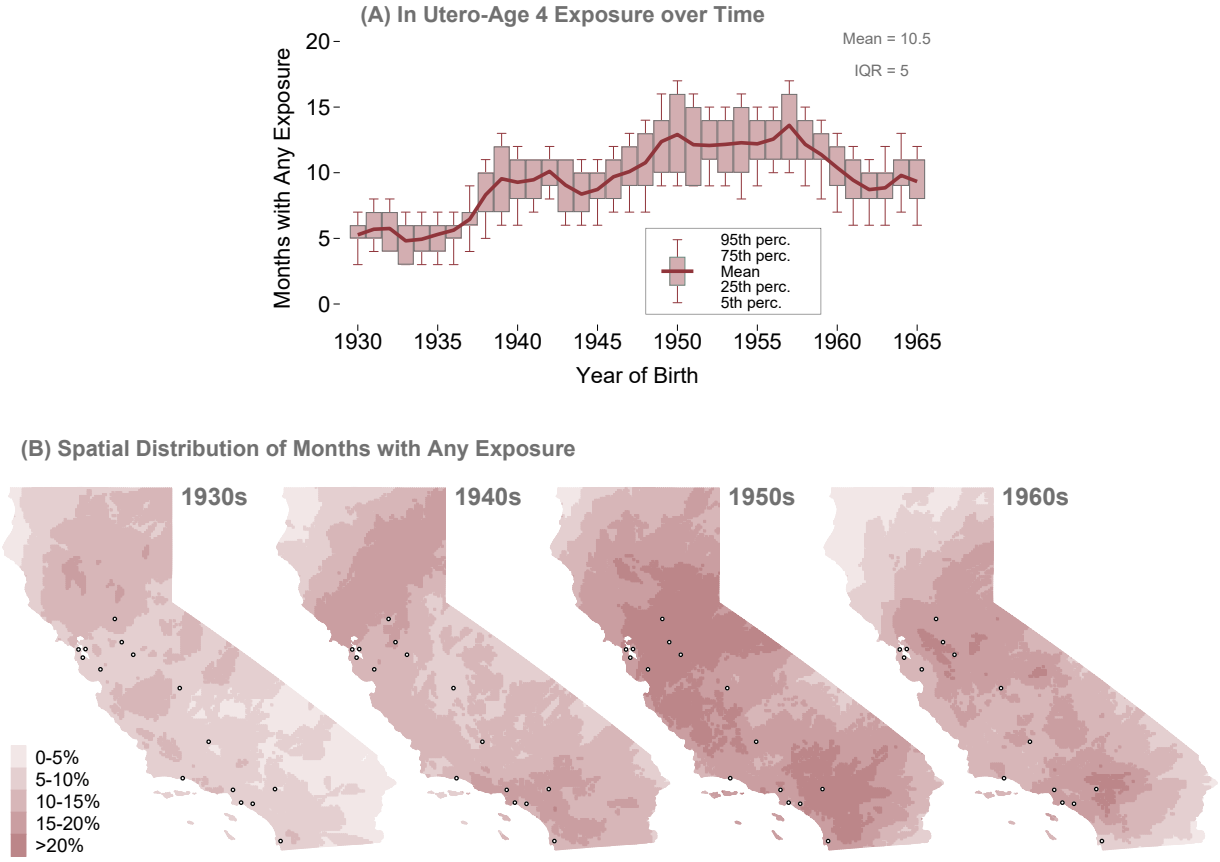
Figure 1 – Descriptive Statistics: There is Substantial Variation in Year-to-Year Acres Burned, but Intensity has Increased Sharply over the Last 30 Years




*Notes:* The purpose of this figure is to display the substantial over-time variation in wildfires during the sample period (Panel A) and the stark increase in wildfire acreage burned over the last thirty years (Panel B). See Section 3.1 for further discussion.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018) and 2021 fire data from `fire.ca.gov/incidents/` (updated through October 11, 2021).

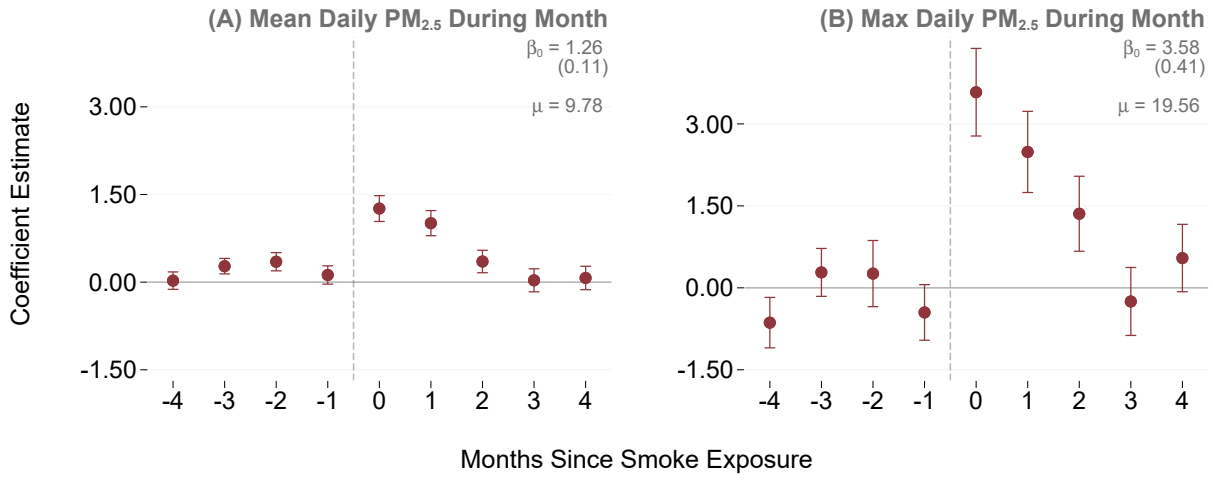
Figure 2 – Identifying Variation: Smoke Exposure Varies Over Time and Geography




*Notes:* The purpose of this figure is to display the identifying variation utilized in our analysis. Panel A displays how the birth-weighted months of smoke plume exposure from an individual's *in utero* period through age 4 ("IU-Age 4") varies across birth cohorts. Within this panel, the mean and various quantiles are plotted for each cohort with the overall mean and interquartile range ("IQR") plotted in the upper-right corner. Panel B displays the spatial variation in smoke exposure according to the percentage of months in a given decade that experienced smoke exposure.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), the Pubic-version Numident (Goldstein et al., 2021).

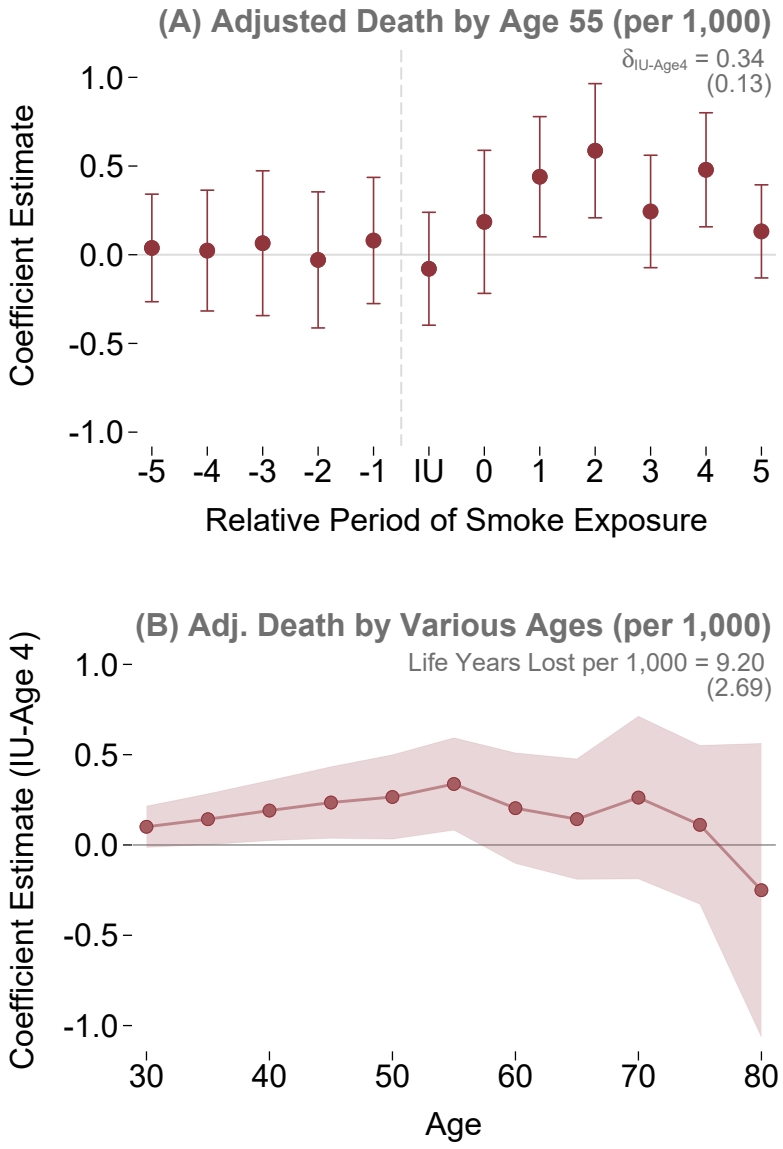
Figure 3 – Validation of Treatment: Modeled Wildfire Smoke Exposure Increases Monthly $PM_{2.5}$




*Notes:* The purpose of this figure is to display the impact of modeled smoke plume coverage on monthly pollution levels. Each point represents a coefficient from the regression detailed by Equation 4 with corresponding 95% confident intervals. The point estimate for the impact on pollution during the month in which the fire occurs ($\beta_0$), along with the standard error and baseline mean are presented in the upper right-hand corner of the graph.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), and Daily Summary Data Files from the EPA Data Mart.

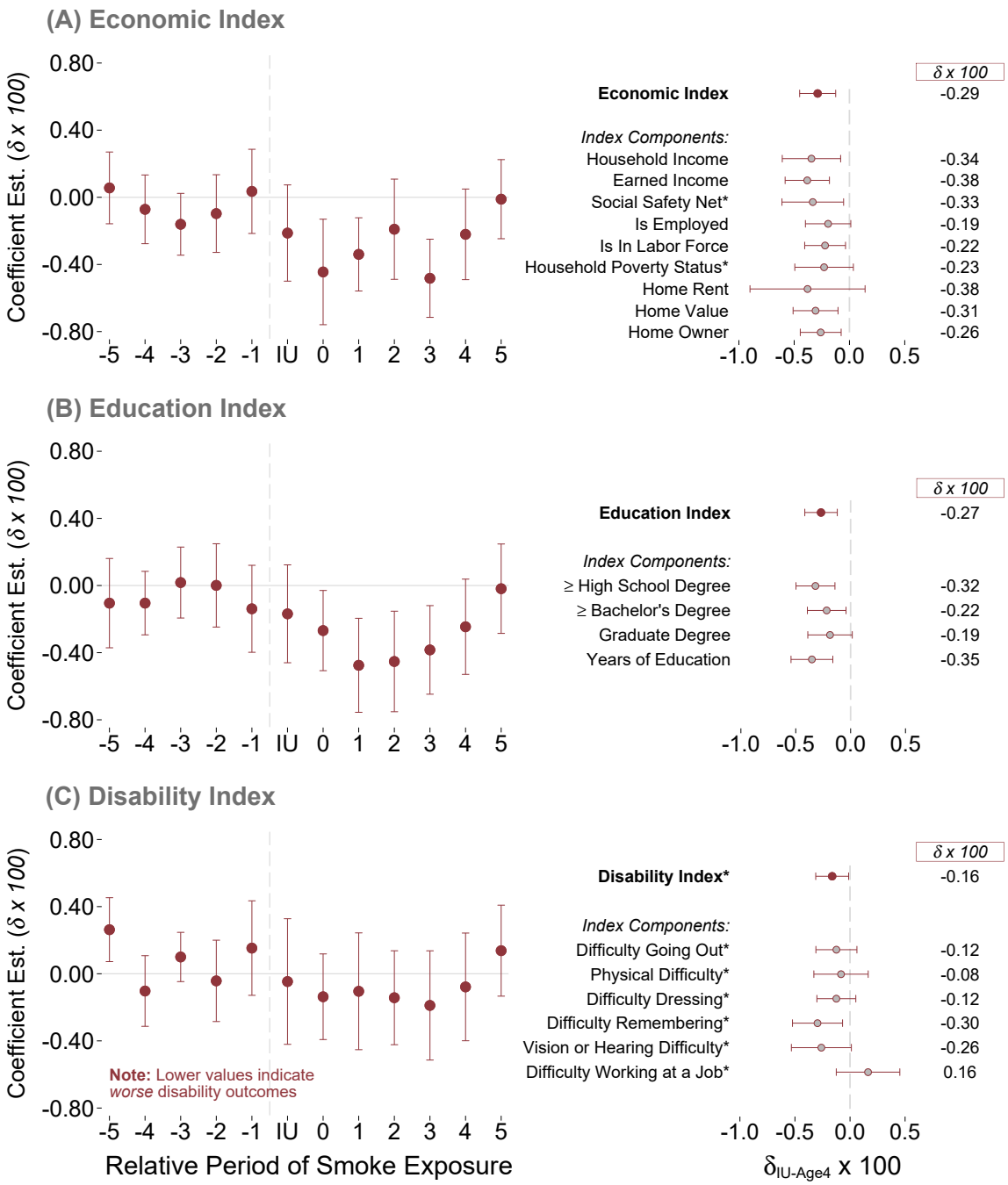
Figure 4 – Main Results: Early-Life Wildfire Exposure Reduces Longevity in Adulthood




*Notes:* The purpose of this figure is to display the impact of wildfire smoke exposure on longevity in adulthood. Within Panel A, each point represents a $\delta_k$ coefficient from the regression detailed by Equation 2 with corresponding 95% confident intervals, representing the marginal effect of an additional month of smoke exposure during the stated period. The upper right-hand corner displays the summary coefficient for an additional month of exposure during the *in utero* period through Age 4 (Equation 3), along with the associated standard error. Panel B displays the summary coefficients from Equation 3 and associated 95% confidence intervals for the effect of an additional month of exposure on at death by various ages (30 through 80).

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), and Restricted Census Numident.

Figure 5 – Main Results: Early-Life Wildfire Exposure Reduces Long-Run Economic and Educational Achievement; Increases Disability

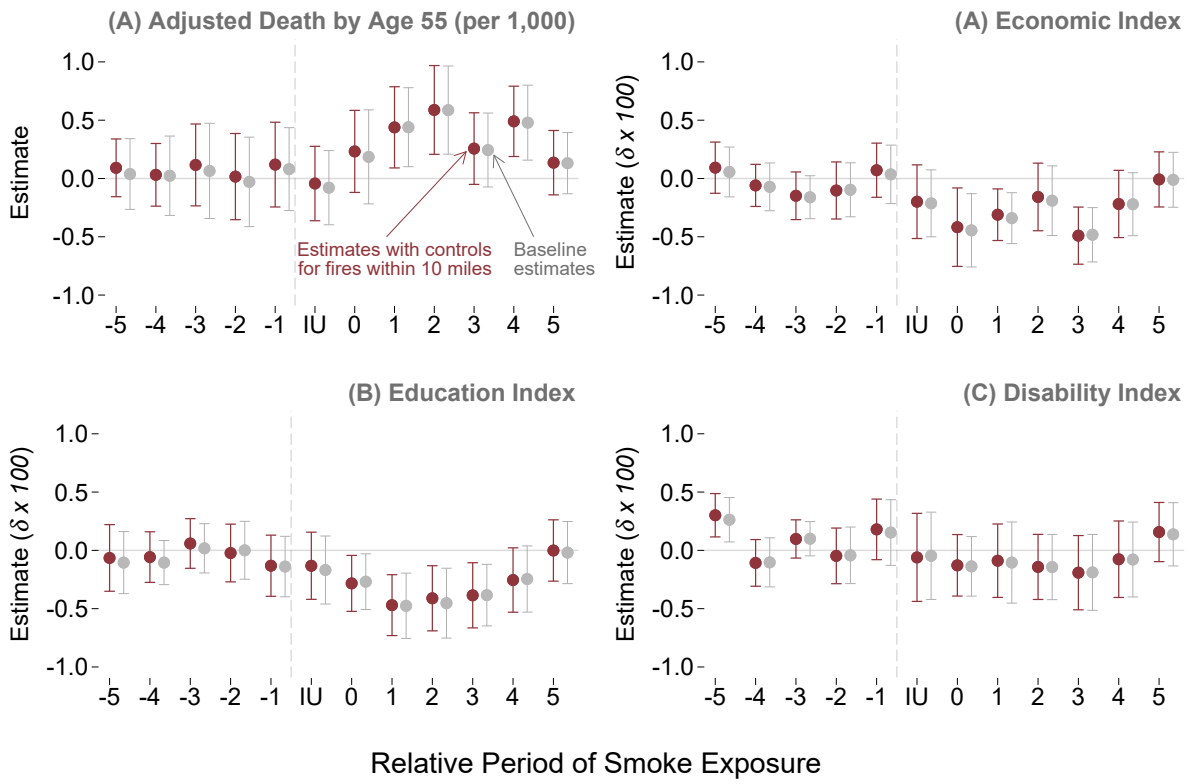



* Indicates that measure has been reverse coded so that higher values correspond with better outcomes (and v.v.).

*Notes:* The purpose of this figure is to display the impact of wildfire smoke exposure on long-run measures of economic achievement, educational attainment, and disability. Within the figure, each panel represents a different index of outcomes with leads-and-lags coefficients (from Equation 2) plotted on the left side of each panel, while summary estimates (from Equation 3) of the effect over the *in utero* through Age 4 period are plotted on the right. These summary coefficients are present for the applicable index and each of its components and are accompanied by the actual coefficient values (in the "$\delta \times 100$" column). All estimates in this figure should be interpreted as the percentage-of-a-standard-deviation change in response to an additional month of smoke exposure during a given period.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, and Restricted American Community Surveys.

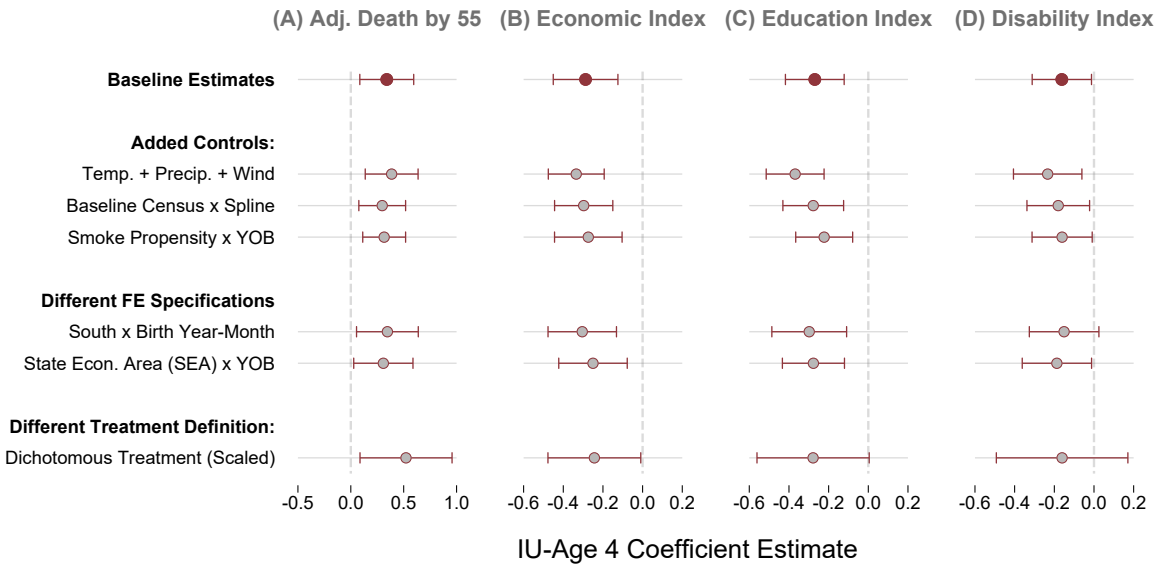
Figure 6 – Robustness: Controlling for Nearby Fires Does Not Affect Estimates, Implying that Smoke Exposure is Driving Effects




*Notes:* The purpose of this figure is to demonstrate that smoke exposure, rather than proximity to wildfires themselves, are driving for estimates. To do so, the figure compares our baseline estimates from our leads-and-lags specification to those when we re-estimate the specification with controls for fires that are within 10 miles of an individual's city of birth. See the text of Section 8.1 for further discussion.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, and Restricted American Community Surveys.

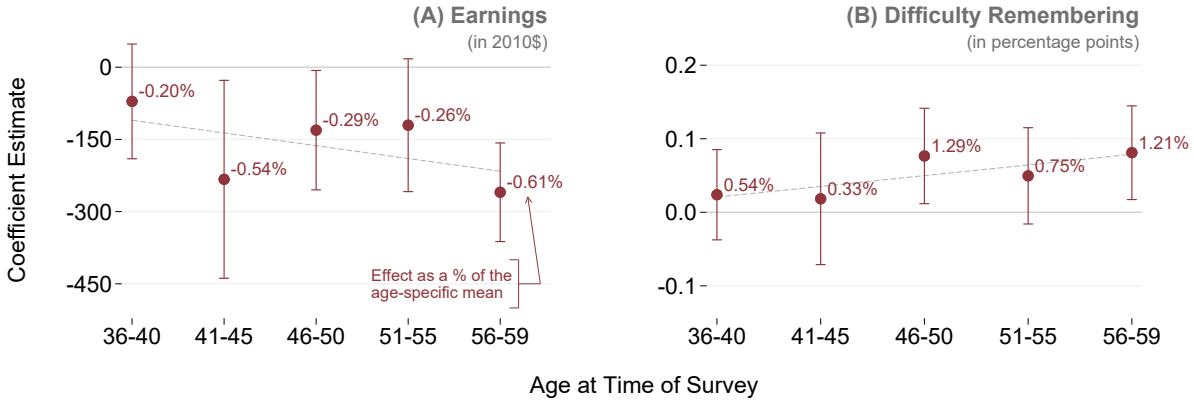
Figure 7 – Robustness to Additional Controls and Specifications




*Notes:* The purpose of this figure is to display the robustness of results to differing specifications. Within the figure, our summary coefficients (Equation 3) for our preferred specification are presented on the top line, with estimates for differing specifications (due to added controls, changes to fixed effect structure, or modifications of functional form) included in the rows below. See Section 8 for further description.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, and Restricted American Community Surveys.

Figure 8 – Supplemental Results: The Effects Early-Life Wildfire Exposure Increase Proportionally with Age for Key Time-Varying Outcomes



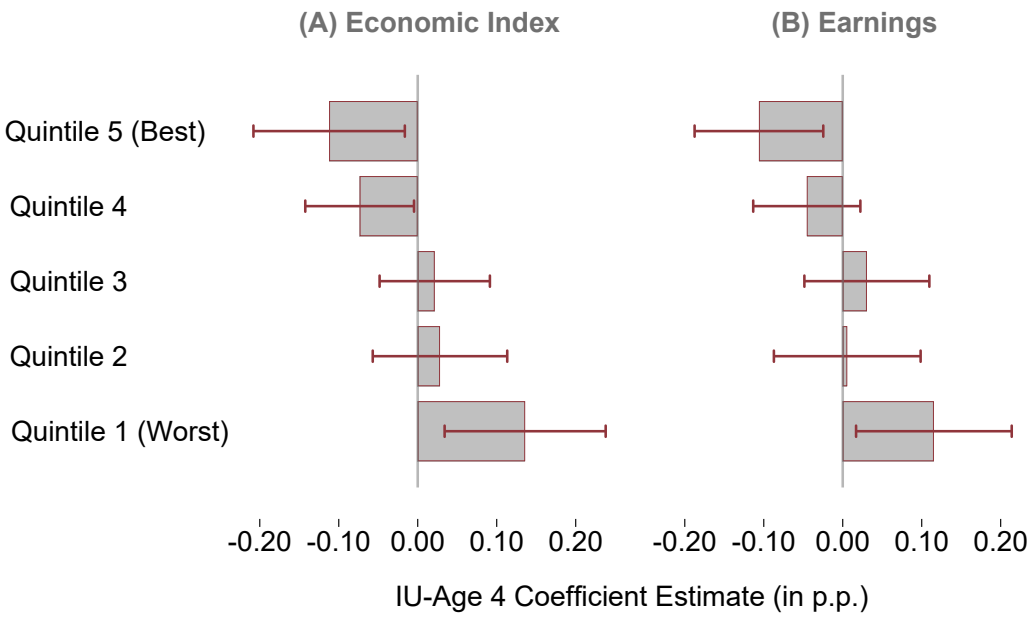

*Notes:* The purpose of this figure is to display how the effects of childhood wildfire smoke exposure evolve as individuals age. Within the figure, each point represents a separate estimate of $\delta_j$ with associated 95% confidence intervals, from the following equation:

$$Y_{cba} = \mu_c + \lambda_{y(b),r(c)} + \alpha_{m(b)} + \psi_{y(b),a} + \sum_{j \in \mathcal{J}} \delta_j \cdot (AgeBin_j \times Smoke_c^{\text{IU-Age 4}}) + \varepsilon_{cba},$$

For further context, the effect as a percent of the age-specific mean is also presented, along with a line of best fit for the coefficients. As demonstrated by the figure, the effects of early life wildfire pollution shocks increase proportionally with age, a finding which is discussed further in Section 9.1.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, and Restricted American Community Surveys.

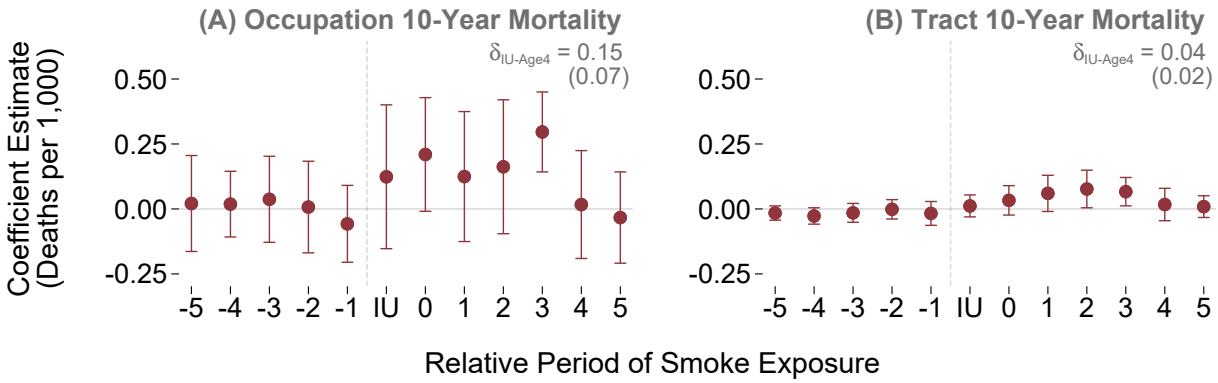
Figure 9 – Supplemental Results: Early-Life Wildfire Exposure Impacts the Entire Distribution of Economic and Earnings Outcomes




*Notes:* The purpose of this figure is to display the distributional impacts of wildfire smoke exposure. Within the figure, each bar represents the coefficient from a separate regression where the outcome is whether or not an individual falls into that quantile (see Section 9.2 for more detail). Estimates are associated with 95% confidence intervals.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, and Restricted American Community Surveys.

Figure 10 – Supplemental Results: Individuals Exposed to Higher Levels of Early-Life Wildfire Exposure Sort into Higher-Mortality Occupations and Neighborhoods




*Notes:* The purpose of this figure is to display the how early childhood wildfire exposure induces individuals to sort into higher-mortality occupations and neighborhoods. See notes to Figure 4 for discussion of graph construction and the text of Section 9.3 for more detail regarding construction of the outcome variable.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, Restricted American Community Surveys, and Restricted Full-count Decennial Censuses.

Table 1 – The Cost and Fiscal Burden of Early-Life Wildfire Smoke Exposure

| **(A) Lost Earnings and Life Years** | | | **(B) Fiscal Burden** | | |
|---|---|---|---|---|---|
| Earnings Loss (Discounted) | $ | 9,892 | Lost Federal Tax Revenue (Discounted) | $ | 1,306 |
| Life Years Lost (Longevity) | $ | 12,558 | Additional Social Safety Net Costs | $ | 482 |
| | | | Additional Incarceration Costs | $ | 99 |
| Loss per Person | $ | 22,450 | Fiscal Cost per Person | $ | 1,887 |
| Loss per 500,000 Persons (in $B) | $ | 11.22 | Fiscal Cost per 500,000 Persons (in $B) | $ | 0.94 |

*Notes:* The purpose of this table is to translate our estimates from Sections 6, 7, and 9 into a measure of the total costs. Within the figure, there are two panels. Panel A summarizes the total per-person costs of lost earnings and lost life years, while Panel B summarizes the total per-person fiscal burden from lost tax revenue and government expenditures. Additionally, both panels scale these results by 500,000 persons, which is the approximate size of a single California birth cohort. See Section 10 for further discussion of how this table was calculated.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, Restricted American Community Surveys, and Restricted Full-count Decennial Censuses.

# Appendix: for Online Publication

# A Data Construction

## A.1 Matching Fire Dates to Meteorologically Similar Days for HYSPLIT Analysis

As discussed in Section 3.1, the available dates for HYSPLIT smoke generation only extend back to 1948, which is in the middle or our sample period. Accordingly, for each fire date, we selected a proxy date that exhibited similar meteorological characteristics as the original day of the fire. Specifically, the following algorithm was performed:

1. Wind vectors—*u*-wind and *v*-wind—were obtained from the NOAA/CIRES/DOE 20th Century Reanalysis dataset ("NCDR data"), which covers the period from 1836-1980, and thus covers days from our original sample as well as days covered by HYSPLIT.[1] Specifically, we obtained data for 3-hour increments of surface wind as well as daily wind vectors for at differing heights above the surface (50 meters, 100 meters, 200 meters, and 500 meters). Days that match along these characteristics are likely to have similar plume dynamics, since both surface and higher-elevation winds determine particulate trajectories. The set of all wind vectors is denoted by $V$, where $v(d)$ is the value of an arbitrary wind vector on a given day $d$.

2. The NCDR data is on a 1-degree latitude by 1-degree longitude grid. To assign more weight to nearby fires, inverse distance weighting was utilized with observations further than 100 miles from the fire centroid receiving no weight. These weights are denoted by $\omega$.

3. Finally, for some fire date $f$, the proxy date $d^*$ is chosen from the set of dates between 1948 and 1980 (the dates of NCDR and HYSPLIT overlap, denoted by $D$) to minimize the distance-weighted Euclidean distance between the set of wind vectors ($V$):

$$d^* = \text{argmin}_{d \in D} \left[ \sum_{v \in V} \omega (v(f) - v(d))^2 \right]^{1/2}$$

4. In order to have consistency across the sample of fires, this procedure is performed for *all* fires, not just those occurring prior to 1948.

## A.2 Place of Birth String Matching Algorithm

As discussed in Section 3.3, the vast majority (over 99%) were hand-matched to a GNIS location or had a perfect text match. Hand-matching was performed for any unique place of birth string that had more than 3 responses in the Public-version Numident or more than 10 responses in the restricted Numident. For the remaining observations, we used a bigram matching method to compare the place of birth string to both (a) names of GNIS locations and (b) POB strings that had already been hand-matched, where the best match was retained. Match similarity was calculated using Jaccard similarity scores; only matches with a score of at least 0.5 were retained. To further

[1] Wind vectors can be utilized to calculate both the angle at which the wind is blowing, as well as wind speed.

validate our algorithmic match, we utilized the place-of-birth-string-to-*county*-of-birth crosswalk from Bailey et al. (2020) and Isen et al. (2017b) (see those papers for more details). Then, using our hand-matched sample, we estimated a predicted probability of correct match based on the Jaccard similarity score and whether the county of the algorithmically matched GNIS location agreed with the county per the Bailey et al. (2020) and Isen et al. (2017b) crosswalk. Matches that had a predicted correct probability of 80% or more were retained in the final sample. As noted in Section 3.3, we matched 99.5% of individuals to a GNIS location.

### A.3 Creation of Adjusted Numident Death Rates

Because the Numident omits potential deaths during the 1930s through mid-1970s, adjustment of the Numident death rates is necessary for our analysis of longevity. To create adjusted rates, the following procedure was implemented:

1. Conditional death rates—e.g., the rate of death by age 75, conditional on surviving to age 14—were calculated using the SSA Cohort Life Tables. These rates are denoted as $r^{ssa}_{bfsd}$ where $b$ denotes the year of birth, $f$ denotes a person's sex, $s$ denotes the "survived-to" age (14 in our example), and $d$ notes the "death-by" age (75 in our example).

2. Similar rates were calculated for all individuals born in the United States from 1930-1969 that are included in the Numident (not just those born in California). For these cohorts, the age at which they received their SSN was used as the "survived-to" age, $s$. These rates are denoted by $r^{num}_{bfsd}$.

3. Since Numident includes a number of individuals with improperly missing deaths, we also utilized the 2000 and 2010 Decennial Full-Count Censuses and the 2005-2019 American Community Surveys to identify individuals that were (a) missing death information in the Numident and (b) did not appear in any of these Census products. Given that an individual should theoretically appear in each Decennial and has roughly a 22.5% chance of appearing in at least 1 American Community Survey, individuals who are not included are at a high risk of having an unrecorded death.[2] Let the measure of missingness for each cohort-sex-SSN age cell be denoted as $\eta_{bfs}$.

4. Then, using this national-level information, an "adjustment factor" to scale cohort death rates is as follows:

$$a_{bfsd} = \frac{r^{ssa}_{bfsd}}{r^{num}_{bfsd} + \eta_{bfs}}.$$

   In words, this adjustment rate scales up the deaths per the Numident (plus the number

[2] While it is possible that there are other reasons for exclusion (e.g., migration out of the United States, homelessness), omitted death is likely the most plausible explanation, particularly for older cohorts.

of individuals likely missing death information) to match national birth cohort-by-sex-by-survival age-by-death age moments per the SSA cohort tables.

5. Finally, this rate is applied to death rates for given cells in our underlying data, such that the adjusted death rate at age $d$ for cohorts in a given birthdate-city of birth-race-sex-SSN age cell will be:

$$r^{adj}_{bcfrsd} = r^{unadj}_{bcfrsd} \times a_{bfsd}.$$

While these rates are used for the primary mortality analyses in this paper, as discussed in Section 6, use of the unadjusted dates results in very similar estimates of life years lost due to smoke exposure.

### A.4 Data Construction for Control Variables and Heterogeneity Analysis

As discussed in Section 3.4, we utilize a variety of datasets to obtain control variables (used in Section 8) as well as variables for supplemental and heterogeneity analysis (Section 9). These additional datasets are described below:

**Aggregate County-Level Data, 1930-1970**. County-level data are obtained from three primary sources: (1) 1930 and 1940 Full-Count Decennial Censuses (Ruggles et al., 2021), (2) County Books from 1944-1970 (U.S. Census Bureau, 2012)—used to provide county averages for post-1940 outcomes—and (3) the 1934 Internal Revenue Service Statistics of Income ("IRS SOI"). We digitized the IRS SOI data to obtain information on family incomes, which is otherwise not available on the county level until the 1950 Census.[3]

**American Hospital Association Data, 1947-1973** Hospital-level data are obtained from historical American Hospital Association data surveys which were transcribed by Finkelstein (2007).[4] We standardized hospital names over time and also mapped city and state location strings to features in the GNIS database with associated geocoordinates. We then generated measures of hospital capacity by calculating the number of beds in general hospitals, pediatric hospitals, and labor and delivery hospitals within the 25 miles of an individual's city and state of birth, which were divided by the number of births within 25 miles (as proxied by the number of individuals in the Numident file). Bed counts were taken from 1947, which is both the earliest year of data and roughly the mid-point for our analysis period.

**Climate Data, 1930-1969**. These data are obtained from three data sources made available by the NOAA: (1) the Global Historical Climatology Network Daily ("GHCND") files, which provides daily temperature variables at the observation station-level; (2) the NOAA-CIRES-DOE Twentieth Century Reanalysis file, which provides daily gridded ($1^\circ$ latitude $\times$ $1^\circ$ longitude) wind direction and speed information; and (3) the University of Delaware Precipitation file (Willmott

[3] The 1940 Census includes only wage income, which is top coded at $5,000 (roughly $78,000 in 2010 dollars).
[4] Many thanks to Amy Finkelstein for generously providing the data.

and Matsuura, 2001), which provides monthly gridded (0.5° latitude × 0.5° longitude) precipitation data. Station-level GHCND temperature data was aggregated to the county level by first creating tract-level temperature variables (using inverse-distance weighting to stations within 100 kilometers of the tract) and then aggregating population-weighted tract-level measures to the county level. Gridded wind and temperature data were aggregated to the county level via a simple average of all grid points within a county or within 1° of the county's border.

# B Additional Results and Robustness

## B.1 Infant Mortality

As discussed in Sections 6 and 8, we perform analyses on infant mortality to compare magnitudes against the reductions in adult longevity and to assess the degree of mortality selection. In the paragraphs that follow, we briefly discuss the data, methods, and regression results for our infant mortality analysis.

### Data

Our primary data for the analysis of infant mortality come from birth and death indexes provided by the California Department of Public Health. The birth index provides full name, date of birth, and county of birth, while the death index provides full name, date of birth, date of death, and county. These data cover our entire sample period of 1930-1969.

To create infant mortality rates, we generated counts of births and deaths for each year-month birth cohort (defined based on date and county of birth) from the underlying vital statistics data. One challenge in using these data is that the death records did not include date of birth for deaths occurring prior to 1940. To adjust for this issue, we matched observations from the birth data using name and county information to construct cohort-level rates for these periods. These rates were then adjusted so that annual *period* infant mortality matched aggregated annual vital statistics tabulations from Bailey et al. (2018). However, our results are robust to omitting observations from the 1930s that were constructed in this manner.

### Methods

We estimate cohort infant mortality using a method that is similar in spirit to our long-run cohort leads-and-lags estimation (Equation 2):

$$IMR_{Cb} = \mu_C + \lambda_{r(C),b} + \sum_{j=-18,-15,...,9} \delta_j Smoke_{C,b+j} + f(X_{Cb}) + \varepsilon_{Cb},$$

where the unit of analysis is a birth cohort born in a county $C$ in year-month $b$. The treatment variable, $SmokeExposure_{C,b+j}$, and its associated leads and lags detail the number of months with any smoke exposure in three-month bins relative to birth. As in our long-term analyses, the coefficients on the leads and lags give us insight into the critical periods of pollution exposure (for $j \in \{-9, -6, ..., 9\}$) and also provide placebo estimates to support our research design (for $j \in \{-18, -15, -12\}$).

We include county-of-birth fixed effects ($\mu_C$) and include region-of-birth $\times$ birth cohort fixed effects ($\lambda_{r(C),b}$) where region of birth is defined as Northern or Southern California. Within our controls, $f(X_{Cb})$, we also include county by birth cohort trends to increase precision. All standard errors for this analysis are clustered at the county level.

**Regression Results**

The results of this analysis are displayed in Appendix Figure A14. Within the figure, each point (with associated 95% confidence intervals) represents the marginal effect of an additional month of smoke exposure within a given three-month bin (e.g., the coefficient at "0" represents the marginal effect in the three months following birth). Based on the leads and lags coefficients, the critical period for exposure appears to be in the three months preceding and following birth (indicated by $\delta_{-3}$ and $\delta_0$, respectively). The average coefficient across these two bins, along with the standard error, is presented in the upper-right corner, while the mean is presented in the lower right for context. We find that an additional month smoke exposure in the last trimester or first three months of life results in 0.36 additional infant deaths per 1,000 births.

To facilitate comparisons between the results discussed in Section 6, we calculate the life years lost due to these additional infant deaths when moving from the 25th to 75th percentile of smoke exposure during the critical period. As discussed in the main text, moving from the 25th to 75th percentile of smoke exposure during the *in utero* through age 4 period results in 5 additional months of wildfire pollution exposure. A similar increase of smoke exposure would result in an increase of 0.43 smoke months during the much shorter critical period for infant mortality. Assuming that the typical infant death would have counterfactually lived to age 75 (roughly the life expectancy at birth for cohorts in our sample), then a $p_{25} \rightarrow p_{75}$ increase in exposure results in 11.7 life years lost per 1,000 persons. For comparison, a single month of exposure—which is close to the *average* level exposure within our sample (0.91 months)—results in 27.0 years of life lost per 1,000. See Section 6 for further discussion of how this compares to reductions in adult longevity.

Finally, in Section 6, we discuss the fact that our adult longevity estimates omit childhood mortality, since they are conditional on obtaining a Social Security Number. (The average age of SSN receipt was 14.4 in our sample.) Accordingly, Appendix Figure A16 adds the infant mortality results into the coefficient estimates from Figure 4, Panel A to provide further context.

## B.2 Comparison of Results with Existing Literature

In this section, we discuss how our findings compare to results from papers that examine the long-run effects of the Clean Air Amendments ("CAAA") of 1970 and 1990, both of which substantially *reduced* particulate emissions. As discussed in Section 7.1, our scaled effects imply that a 1 $\mu g/m^3$ increase in *annual* $PM_{2.5}$ results in a 2.5% decrease in earned income.[5] As noted in the main text, this estimate is slightly smaller than the most recent estimates of the long-run impact of air pollution on early adult wages by Colmer et al. (2022), who leverage variation from the 1990 CAAA and find that a 1 $\mu g/m^3$ decrease in annual $PM_{2.5}$ results in a 2.8% to 4.3% increase in wages.

In addition to the results from Colmer et al. (2022), there are two other papers—Isen et al.

[5] This is calculated by taking our estimated decrease in earned income from Appendix Figure A17A and dividing it by the average earned income from Appendix Table A1 to obtain the percentage effect of an additional month of wildfire exposure of 0.4% (\$178/\$44,087). This is then scaled based on the sum of the post-period estimates from Figure 3 (less the sum of pre-period estimates), divided by 12 to obtain an annual pollution impact ($\frac{0.4\%}{1.95/12} = 2.5\%$).

(2017b) and Voorheis (2017)—that merit comparison. Both of these papers utilization variation from the *1970* CAAA and find that, in response to a reduction of 10 $\mu g/m^3$ in TSPs, there was an increase of approximately 1.4% in wage earnings by ages 28-32 (Isen et al., 2017b) and a 0.7 to 0.9 to percentage point increase in high school completion (Voorheis, 2017). These papers measure effects primarily in TSPs, which do not translate directly to changes in fine particulate matter pollution ($PM_{2.5}$), which we measure in this paper, and which is generally seen as a better indicator of the potential of air pollution to harm human health.[6] However, based on comparisons for monitors that have contemporaneous measurements of both TSPs and $PM_{2.5}$, we note that a 1 $\mu g/m^3$ increase in $PM_{2.5}$ is associated with a 4-5 $\mu g/m^3$ increase in TSP. Accordingly, the estimates from the CAAA papers stem from a reported decrease that is equivalent to 2-2.5 $\mu g/m^3$ of $PM_{2.5}$. Translating these estimates implies that a 1 $\mu g/m^3$ decrease in annual $PM_{2.5}$ results in a 0.6-0.7% increase in wages and a 0.3-0.5 percentage-point increase in high school completion. Therefore, our scaled estimates—which imply that a 1 $\mu g/m^3$ *increase* in annual $PM_{2.5}$ results in a 2.5% decrease in earned income and a 0.5 p.p. decrease in high school completion—are larger than Isen et al. (2017b) but roughly the same as Voorheis (2017). However, as discussed below, while these studies provide compelling causal evidence of the reduced-form impact of the CAAA, comparing their instrumental variable effects (noted above) to the results that we find is challenging for several reasons.

The first set of challenges relates to contextual differences for these papers that may result in meaningfully different results. We begin by noting that the pollution sources that we study are compositionally different. While there is no direct evidence (absent this paper) that childhood wildfire pollution shocks result in worse long-term outcomes, there is a growing body of evidence focusing on contemporaneous outcomes that suggest wildfire pollution could cause larger adverse impacts than an equivalent amount of pollution from other sources. See Section 2 for further discussion. Additionally, as noted in Colmer et al. (2022), the 1970 CAAA did not focus on finer particulates, which are very prevalent in wildfire smoke and are more harmful to development than large particulate pollution. Accordingly, the actual reduction in $PM_{2.5}$ from the 1970 CAAA may be substantially less than our crude conversion from TSPs to fine particulates implies, lessening the overall impacts.

Next, as detailed in Section 9.1, our effects grow substantially over time: the effect for our oldest age group is three times as large (in percentage terms) as the youngest. If we had instead estimated effects using only our youngest age group, we would find percentage wage reductions of approximately to 1.2% per unit of $PM_{2.5}$, which are much closer to those found by Isen et al. (2017b). These age effects also help to explain why we find similar effects on educational outcomes as Voorheis (2017) (since education is largely determined by age 30 and invariant thereafter), whereas we find larger effects than Isen et al. (2017b) (because we focus on later ages). It is also worth noting that, as discussed in Section 9.2, wildfire smoke appears to have different effects across the outcome distribution than the CAAA did. Whereas we find that childhood wildfire smoke exposure shifts

[6] Measurements of $PM_{2.5}$ were not available for the study periods in these papers.

the entire earnings distribution to the left, Isen et al. (2017b) finds that the left tail is compressed into the middle of the distribution.

The next set of challenges concerns the measured pollution effects of the CAAA and the excludability of the Clean Air Act Amendments as an instrument. First, as acknowledged by Isen et al. (2017b) and Voorheis (2017), the CAAA induced decreases in economic activity (Greenstone, 2002; Walker, 2013). While these economic shocks are widely regarded as small, they could meaningfully down-bias the long-run impacts of air pollution, especially if they are concentrated among vulnerable populations. Next, since the completion of Isen et al. (2017b) and Voorheis (2017), there have been several studies that critically examine the role of "gaming" pollution regulations. One such paper (Gibson, 2019) notes that pollution reductions from later Clean Air Act Amendments are limited to industrial activity around non-attainment monitors, suggesting that the average county-wide pollution reductions stemming from the CAAA of 1970 could be meaningfully lower than previously thought. Gibson (2019) additionally finds that regulation from later Clean Air Act Amendments led to increases in water pollution due to the substitution of firms to different pollution types. Given that industrial water pollution has been shown to negatively impact infant health (e.g., Flynn and Marcus, 2021), this will further offset the benefits found by the CAAA. Lastly, recent research has shown that the monitors readings themselves are subject to manipulation behavior, and so observed reductions in pollution may not reflect true improvements in air quality. For instance, Zou (2021) finds that firms displace their emissions from days where monitors are active to days in which monitoring does not occur. This displacement results in a 7% reduction in *measured* pollution (relative to baseline levels) for counties near regulatory thresholds. Additionally, Grainger et al. (2017) find that regulators in counties near or above regulatory thresholds strategically place new monitors in "clean" areas to reduce the average monitored pollution in these counties. This "gaming" of regulations may indicate why the pollution-reduction impacts of the 1970 CAAA are nearly 50% of their original magnitude when restricting to a panel of balanced monitors which are therefore not affected by strategic placement or retirement (Cropper et al., 2022).

Taken together, these factors make it highly difficult to make reasonable comparisons between our findings in this paper and other papers which utilize variation from the Clean Air Act Amendments to identify long-run effects. Accordingly, further research is needed to understand the causal chain that leads from early-life air pollution exposure in later life outcomes.

### B.3 Bounding Mortality Selection

As discussed in Section 8, we implement a procedure to bound the impact of mortality selection on our estimates. While that exercise is described generally in the text, we discuss it in detail here. We begin by estimating age-of-exposure effects for death by age $a$ using the techniques described in the text.

Explicitly, for $a \in \{30...59\}$, we estimate:

$$M^a_{cb} = \mu_c + \lambda_{y(b),r(c)} + \sum_{k=-5}^{5} \delta^a_k Smoke_{c,b+k} + \varepsilon_{cb},$$

where $M^a$ is the cumulative mortality by age $a$, conditional on surviving past childhood.

Using the estimates recovered from the previous step, we predict the fraction of individuals that would have died by the age we observe at a given survey $(r^a)$ for each cohort (defined on the city $\times$ year-month of birth level):

$$r^a_{cb} = \sum_{k=-5}^{5} \hat{\delta}^a_k \times Smoke_{c,b+k}$$

This exercise also incorporates the infant mortality estimates from Appendix Section B.1, which are calculated by multiplying the number of smoke months within the 3 months on either side of birth by the summary coefficient over that period (0.36 per 1,000).

We then use the death-by-age rate to construct an upper and lower bound. To do so, we create new cohort-level outcome measures by combining the observed average of the outcome $(\bar{y}_{cba})$—weighted by the observed individuals in that cohort $(n_{cba})$—with an extreme value $(g)$ that is weighted by the estimated number of individuals who died before the survey date (calculated using $r^a_{cb}$):

$$y^g_{cba} = \frac{(\bar{y}_{cba} \times n_{cba}) + (g \times \frac{r^a_{cb}}{1-r^a_{cb}} \times n_{cba})}{n_{cba} \times \frac{1}{1-r^a_{cb}}}.$$

In the equation above, the term $\frac{r^a_{cb}}{1-r^a_{cb}} \times n_{cba}$ captures the number of individuals who are predicted to have died before the survey, while the denominator $\left(n_{cba} \times \frac{1}{1-r^a_{cb}}\right)$ equals the number of individuals who are observed, plus those who have died. The value of $g$ that is selected is equal to either the 99th in-sample percentile of the outcome distribution (for the upper bound) or the 1st in-sample percentile (for the lower bound).

Finally, using these revised outcome measures, we re-estimate the following equation for the upper and lower values of $g$ to obtain our bounded estimates:

$$y^g_{cba} = \mu_c + \lambda_{y(b),r(c)} + \alpha_{m(b)} + \psi_{y(b),a} + \sum_{k=-5}^{5} \delta_k Smoke_{c,b+k} + \varepsilon_{cba}.$$

The results of this procedure are presented in Appendix Figure A21. As displayed by the figure, the bounded estimates are highly similar to our original estimates.[7] We further note that, of the two bounds presented in the figure, the lower bound is likely more representative of the "true"

[7] A careful review of the figure will show that, while the upper and lower bounds for our economic and education outcomes are generally symmetrical around the baseline estimate, this is not the case for the disability index. This is because the best possible outcome for the disability index is to have no disabilities at all, which is the case for 84% of our sample. Thus, the upper bound is essentially identical to the baseline and only the lower bound differs.

estimate, because individuals who are in poor health (and thus are more likely to die before we can observe them) also tend to have poor outcomes in other areas as well.

### B.4 Migration-Related Measurement Error

We perform exercises to assess the impact of measurement error on our estimates, particularly those in the leads-and-lags specification detailed by Equation 2. While this procedure is discussed generally in Section 8, we describe it in greater detail here. We begin by utilizing the 1940 Census 100% Sample. These data are used because they ask the entire population what county they lived in on April 1, 1935, and thus allow us to accurately assign both county of birth and county of residence—and their corresponding smoke exposure measures—for 5-year-olds in our sample.[8] Additionally, the sample size allows us to restrict to children aged 5 that were also born in California.

Using this sample of children, we then simulate smoke exposure by fixing a child's birth place and residence at age 5 but assigning smoke as-if that child was born in different periods in our sample. (This exercise implicitly assumes that the migratory behavior of children in 1940 is applicable to all cohorts.) Using the measures of smoke based on place of birth and place of age 5 residence, we then construct "true" measures at each age, where the correct measure of smoke exposure in the *in utero* period and the first year of life (age 0) is equal to the county-of-birth measure. For other ages ($k$), we calculate the "true" smoke measure as:

$$TrueSmoke_k = \frac{\left[Smoke_{birth} \times (5-k)\right] + \left[Smoke_{res} \times k\right]}{5} \quad \text{for } k \leq 5,$$

where $Smoke_{birth}$ and $Smoke_{res}$ represent smoke measures based on an individual's county of birth and residence at age 5, respectively. Note that this exercise that effectively assumes equal likelihood of migration in any given year. Using this true smoke measure, we can calculate the amount of bias under the assumption that the measurement error is classical—an assumption that we revisit below.

For each age, we calculate the measurement error as $u_k = Smoke_{birth} - TrueSmoke_k$ to represent the difference in exposure calculated based on place of birth versus our best estimate of the true exposure calculated above. We then compute the variance of the measurement error ($\sigma^2_{u,k}$) and the "true" age-specific smoke measure ($\sigma^2_{smoke,k}$) to calculate the measurement-error adjusted estimate for each exposure age:

$$\delta_k^{adj} = \frac{\delta_k}{1 - \dfrac{\sigma_u^2}{\sigma^2_{smoke,k} + \sigma_u^2}}.$$

As displayed in Appendix Figure A22 these adjusted estimates are almost identical to our baseline estimates. We conclude this section by briefly revisiting our assumption of classical measurement

[8] Children moving out of state were assumed to have no smoke exposure.

error. While we can't validate all the conditions necessary for this assumption, we do note that, within our data, $\frac{1}{n}\sum_i u_{k,i} \approx 0$ and $\frac{1}{n}\sum_i TrueSmoke_{k,i} \times u_{k,i} \approx 0$. Accordingly, the use of classical measurement error assumptions appears to be appropriate.

### B.5 Other Supplemental Results

In this appendix section, we discuss further supplemental analyses that were only briefly referenced in the text. Specifically, we consider: (1) the degree to which our main results—particular our results concerning mortality—could be explained instead by local economic conditions; and (2) the degree to which there is treatment effect heterogeneity across individual- and place-specific characteristics.

#### The Effects of Wildfire Smoke Do Not Appear to Be Primarily Transmitted through Local Economic Shocks

As demonstrated by Borgschulte et al. (2020), wildfire smoke has the ability to affect labor market conditions, particularly for older workers. Accordingly, it is possible that the impacts of wildfire smoke on later life outcomes are due, in part, to degraded local economic conditions which could then lead to lower levels of childhood investment. To investigate this possibility, we utilize county-level data from 1930-1970 Decennial Censuses and 1934 IRS Statements of Income Data (see Appendix Section A.4) to estimate the following specification:

$$\Delta Y_{Ct} = \alpha + \tau \cdot Smoke_C^{t',t} + \lambda_t + \varepsilon_{Ct}, \tag{B1}$$

where the unit of analysis is a county, $C$, observed at time $t$. The outcome variable is a long-difference (typically of a decade) between time $t$ and the previous period, $t'$. The use of a long-difference specification effectively nets out any time-invariant characteristics, while the time fixed effect, $\lambda_t$, controls for changes common to all areas of California. The treatment variable, $Smoke_C^{t',t}$ represents the number of months with any smoke between time $t'$ and $t$. In addition to our interest in local economic conditions, we also utilize these data to assess the impact of wildfire air pollution on net migration, which we define as:

$$Migrate_{Ct} = \frac{Pop_{Ct} - Pop_{Ct'} - Births_C^{t',t} + Deaths_C^{t',t}}{Pop_{Ct'}}$$

The results from this equation are displayed in Appendix Table A4. Within the table, which presents the marginal effect of an additional smoke month during the period, we do not find any statistically significant results. We also note that results are economically small, suggesting that contemporaneous economic do not play a large role in the long-run effects we find and do not induce migratory behavior. However, to provide further context to these results, we relate our (statistically insignificant) effect on family income to estimates from Aizer et al. (2016), who estimate the impact of large ($\sim$25%) increases in family income on children's longevity. To do so, we calculate

the $p_{25} \rightarrow p_{75}$ effect on income as a percent of the underlying mean (0.4%) and use this estimate to scale the Aizer et al. (2016) finding of a 1-year increase in life expectancy. This procedure yields an estimate of approximately 16 life years lost per thousand individuals $\left(\frac{0.4\%}{25\%} \times 1 \text{ life year} \times 1000\right)$, or 35% of our estimate. We should further note that this figure is likely an upper bound of the degree to which income shocks could affect longevity, because the effects that Aizer et al. (2016) estimate are within the context of very poor mothers with dependent children. Accordingly, the returns on income they find are likely much larger than would be true for the population study.[9]

**The Effects of Wildfire Smoke Have Little Heterogeneity Across Groups**

Next, we consider how the effects of wildfire smoke varies across groups. Specifically, for each group, we estimate:

$$Y_{cba} = \mu_c + \lambda_{y(b),r(c)} + \alpha_{m(b)} + \psi_{y(b),a} + \delta_0 \cdot Smoke_c^{\text{IU-Age 4}} + \delta_1 \cdot \left[Smoke_c^{\text{IU-Age 4}} \times \mathbf{I}(g(i) = g)\right] + \varepsilon_{cba}, \quad \text{(B2)}$$

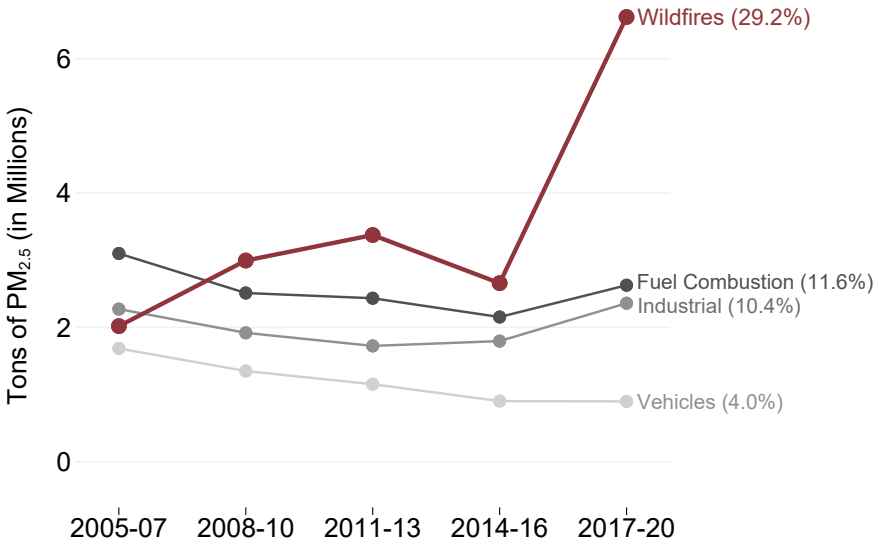
where the indicator variable, $\mathbf{I}(g(i) = g)$, is equal to one if an individual is in the interaction group and zero otherwise. Accordingly, the coefficient on the interaction term, $\delta_1$, captures the additional effect on a month of wildfire pollution relative to the omitted group.

We present results from this equation in Appendix Figure A26. Within the figure, there are two sets of estimates presented for each regression, the baseline effect $(\delta_0)$ and the *combined* effect $(\delta_0 + \delta_1)$ which captures the total effect on the interaction category. Both the baseline effect and combined effect are presented with their 95% confidence intervals. Additionally, estimate pairs that are statistically distinct from each other—i.e., where the interaction term, $\delta_1$, achieves statistical significance—are denoted with one or more "*" indicating the level of that significance.

As demonstrated by the figure, there is very little heterogeneity among groups: only three of the twenty-four regressions have a statistically significant effect. Accordingly, we conclude that wildfire smoke generally has homogeneous effects across a wide variety of groups. See Appendix Section A.4 for further discussion on how these heterogeneity groups were constructed.

[9] However, we utilize these estimates because they are, to the best of our knowledge, the only causal estimates of an early-life income shock on longevity.

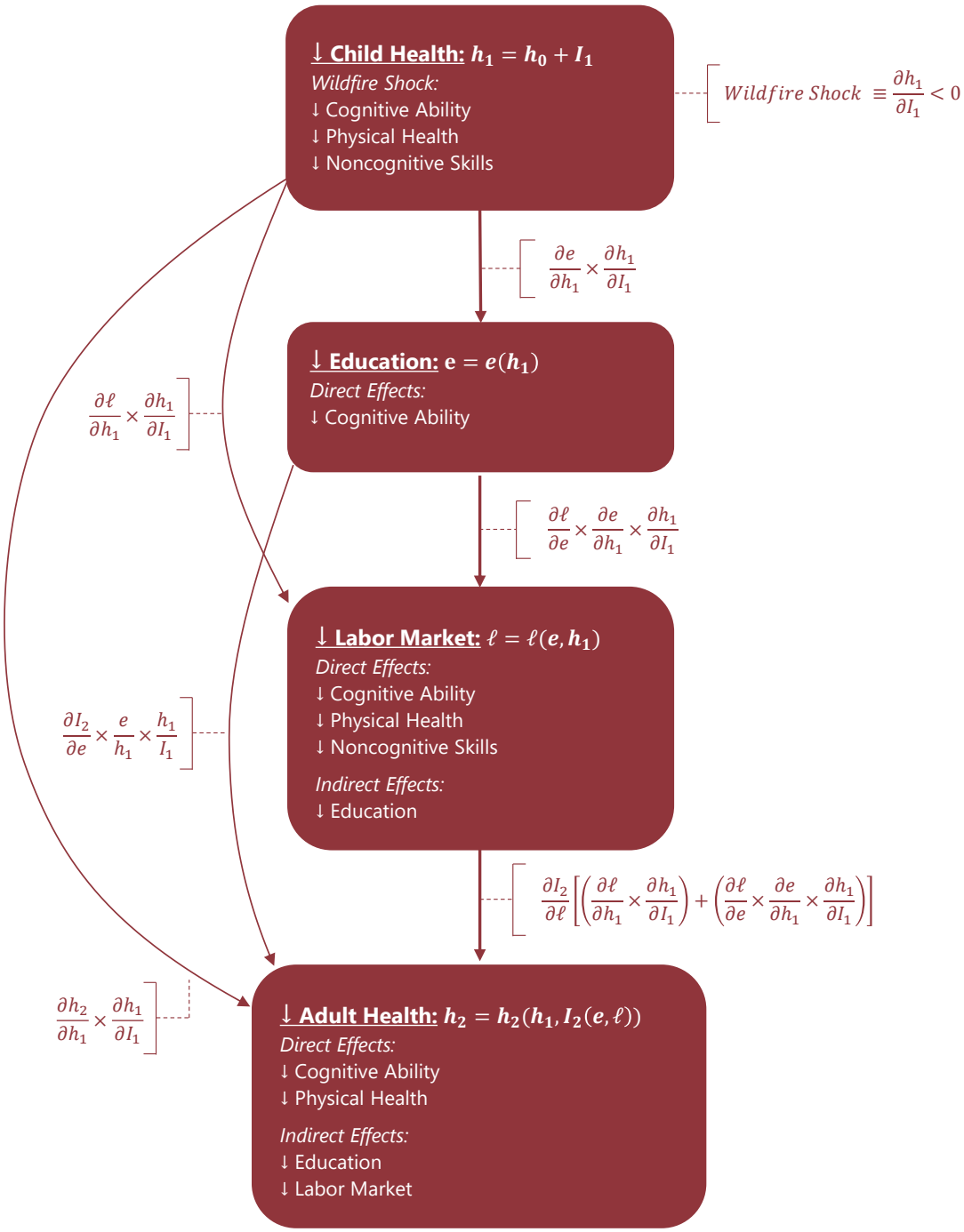
Figure A1 – Motivating Figure: Wildfires are a Growing Source of $PM_{2.5}$ Pollution




*Notes:* The purpose of this figure is to detail trends in the share of fine particulate matter ($PM_{2.5}$) attributable to wildfires and other major sources tracked by the Environmental Protection Agency ("EPA"). The connected points represent trends in total tons of pollution emitted while the percentages to the right of the graph represent the share of $PM_{2.5}$ attributable to each source in the most recent period. Percentages do not add up to 100% due to omission of the EPA's "miscellaneous" category. Years are grouped into bins because the EPA only includes certain pollutant measures as multi-year averages within these data.

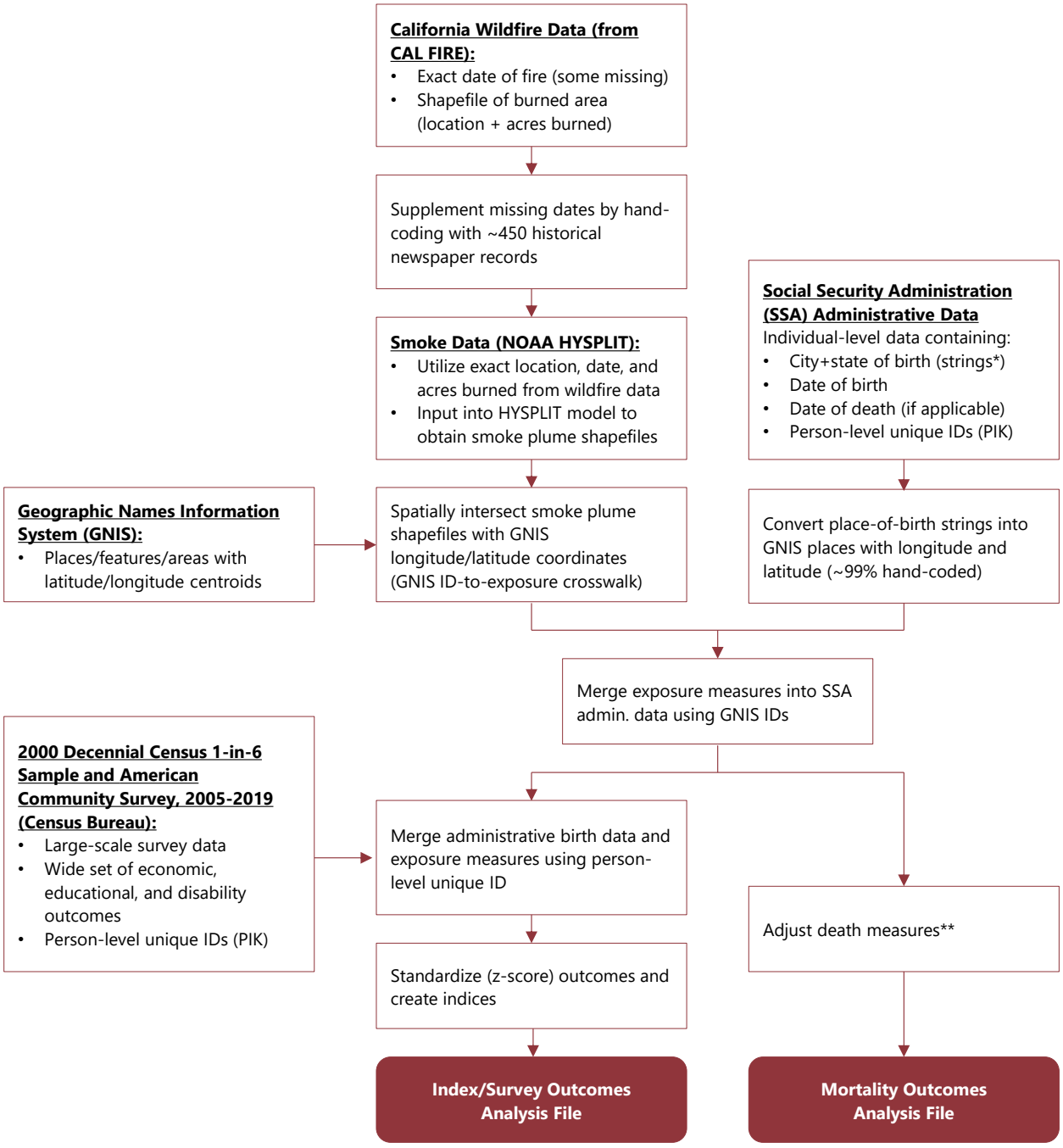
*Source:* Author calculations using National Emissions Inventory Data from the Environmental Protection Agency.

Figure A2 – Illustration of Conceptual Framework: Wildfire Smoke Could Affect Long-Run Health through Multiple Channels




*Notes:* The purpose of this figure is to illustrate the channels by which wildfire smoke exposure during childhood could affect long-run outcomes. Each box in the figure represents a different part of the life cycle, as discussed in Section 2, with the text inside the box describing the ways in which childhood pollution could affect outcomes during that stage. The red text associated with each arrow (and the first box) represent the applicable components of the mathematical model described in Section 2—i.e., they describe how the effects are transmitted from one stage to the next.

Figure A3 – Graphical Illustration of Data Construction for Primary Data Sets

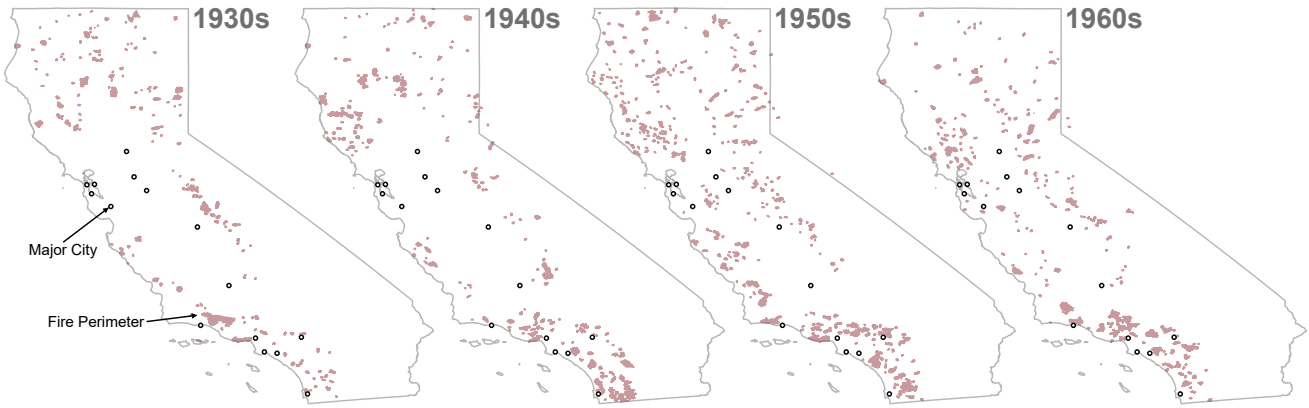



*Notes:* The purpose of this figure is to illustrate the process by which final datasets constructed, which is further discussed in Section 3. Note that this flow chart is not intended to account for all data sources constructed in this paper, only the primary analysis data sets.

* As discussed in Section 3.3, the place of birth data in the administrative SSA data (the Numident file) has not been standardized and is instead presented as a string variable entailing what was originally written on an individual's Social Security application (e.g., an individual born in San Francisco, California might write their city of birth as "San Francisco," "San Fran," or simply "SF"). Accordingly, it was necessary to convert into standardized GNIS places or features.

** The SSA Numident is a nearly comprehensive source of deaths occurring after the mid-1970s but may be missing deaths from earlier years. As further discussed in Section 3.3 and Appendix Section A, we therefore adjusted death rates to account for Numident's propensity to undercount these early deaths.

Figure A4 – Descriptive Statistics: Location of Fires




*Notes:* The purpose of this figure is to display spatial trends in wildfire occurrence. Wildfires occur across the state, although the exact location where they occur varies substantially from year to year.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018).

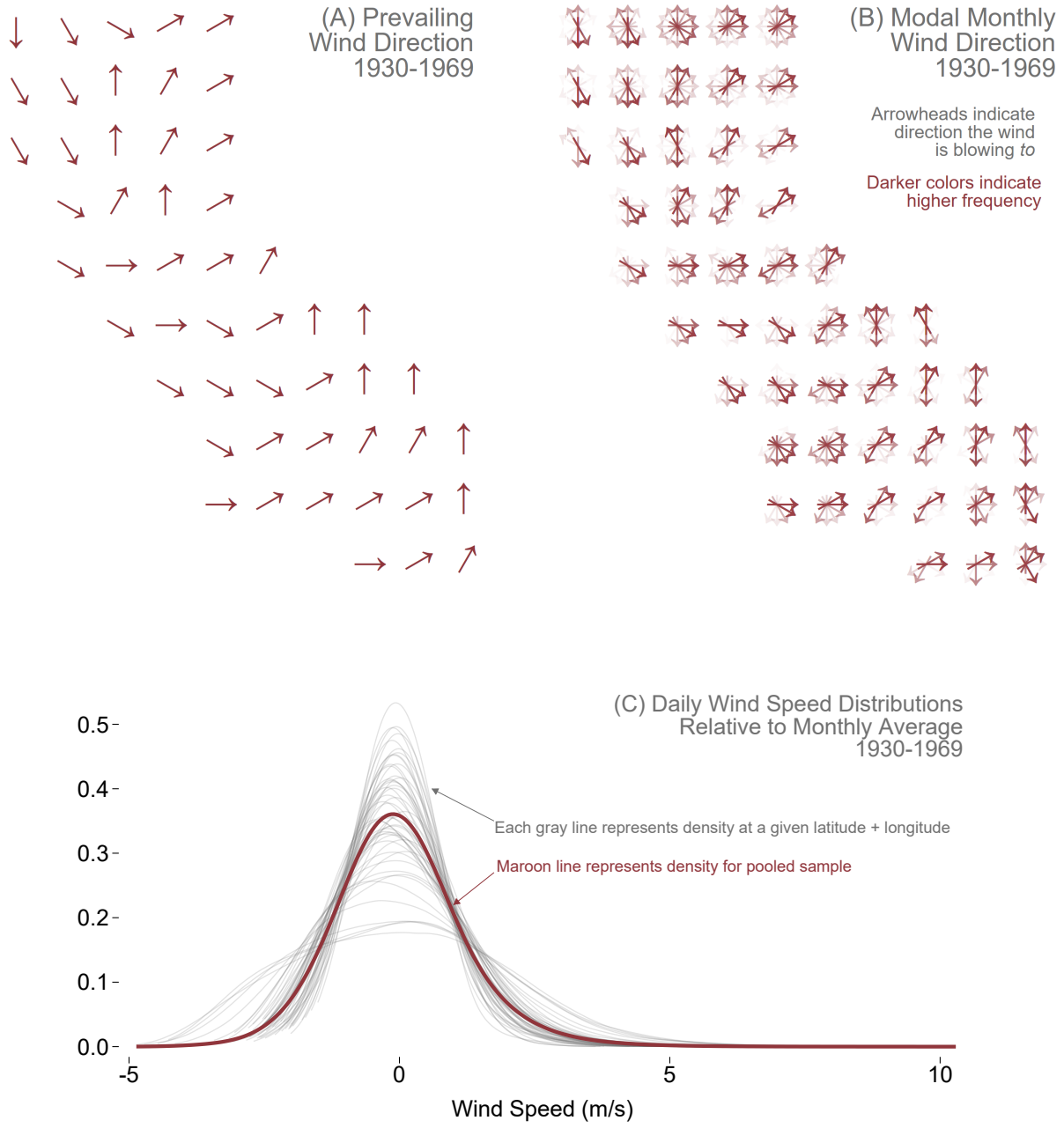
Figure A5 – Illustration of Modeled Smoke Plumes Under Different Conditions




*Notes:* The purpose of this figure is to illustrate different smoke plumes generated by HYSPLIT under different conditions. Panel A illustrates a smoke plume from a hypothetical large (10,000-acre) fire on a day when there are strong winds (around the 90$^{th}$ percentile of wind speed). Under such conditions, there is a substantial amount of smoke generated, and it can travel for several hundred miles. Panel B illustrates a smoke plume from an alternate small (1,000-acre) fire in the same location and under the same wind conditions. In contrast with Panel A, there is a substantially lower amount of smoke predicted. Finally, Panel C illustrates a large fire under calmer wind conditions (around the 10$^{th}$ percentile of wind speed). While there is a substantial amount of smoke generated, as is the case in Panel A, it is spatially distributed in a diffuse mass surrounding the fire. These figures underscore the importance of incorporating both fire size and wind speed into the modeling of wildfire pollution.

*Source:* Author simulations using HYSPLIT model output (Stein et al., 2015).

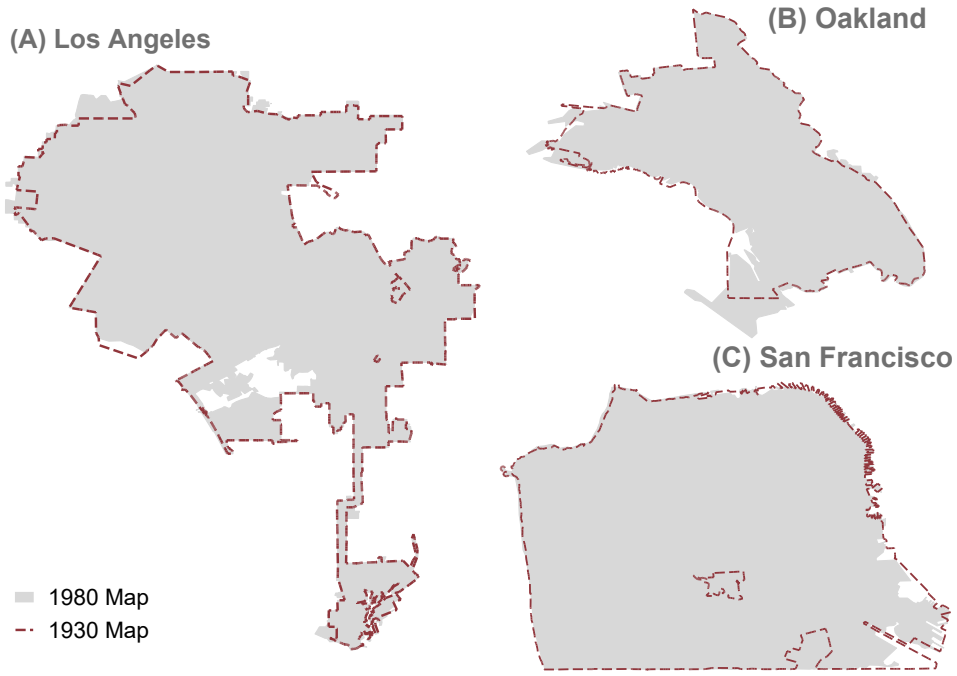
Figure A6 – Descriptive Statistics: Changes in Wind Direction and Intensity Provide Plausibly Exogenous Variation




*Notes:* The purpose of this figure is to illustrate the variation in wind direction and speed. Within the figure, Panel A displays the prevailing wind direction from 1930-1969 (i.e., the modal wind direction calculated over the entire period) for various points of longitude and latitude in California (state boundary lines have been omitted for visual clarity). Panel B performs a similar exercise, except that the modal direction is calculated for each year-month and plotted in the figure using semi-transparent arrows, so that darker arrows indicate more frequent modal directions. Panel C presents the kernel densities of wind speeds for California (maroon line) and for each of the longitude/latitude points plotted in Panels A and B (semi-transparent gray lines). All speeds are relative to the average speed at that point in that month (calculated over the 1930-1969 period) to emphasize variation in average wind speeds. All panels restrict observations to July through November, as that is when the overwhelming majority of fires occur (and therefore when wind direction is relevant) – see Appendix Figure A12 for more detail.

*Source:* Author calculations using the NOAA-CIRES-DOE Twentieth Century Reanalysis File.

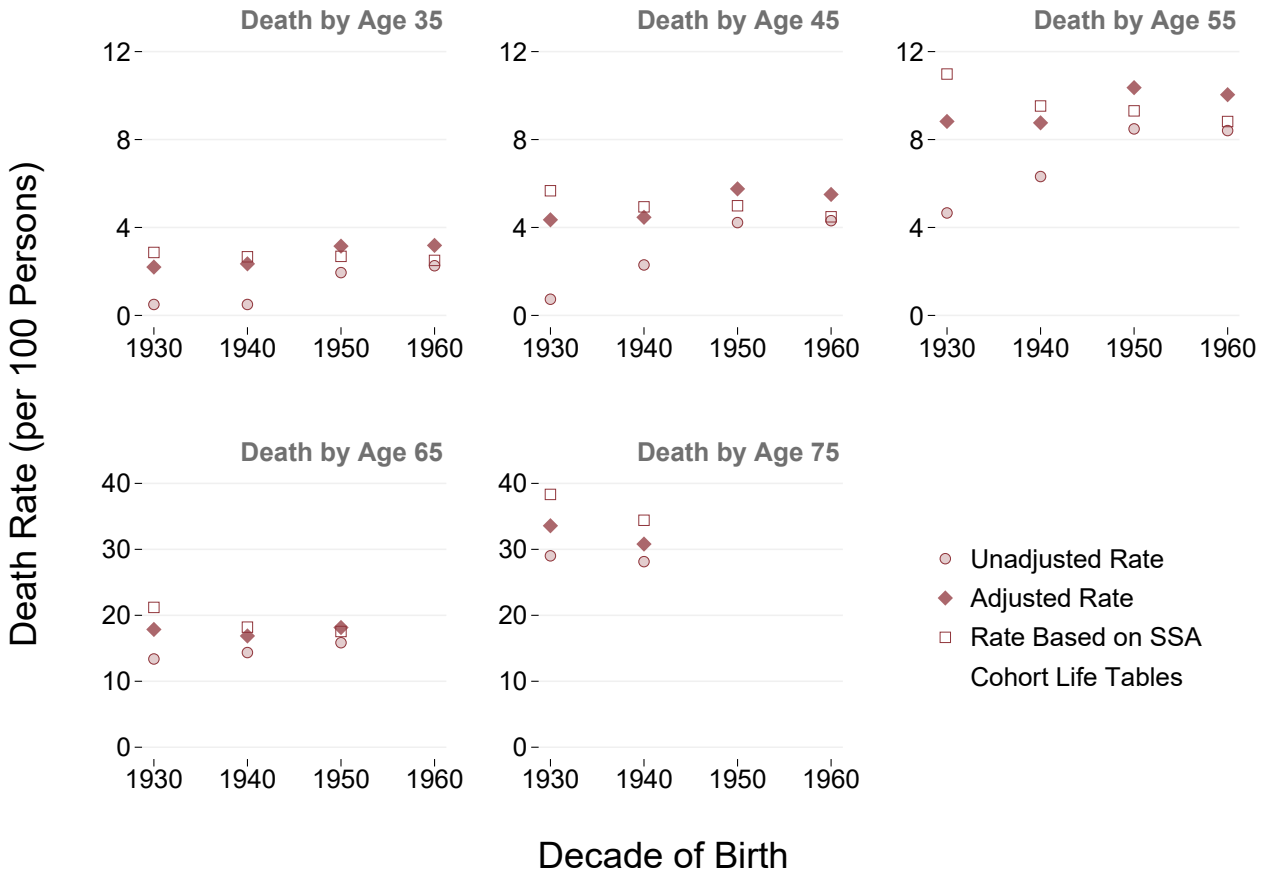
Figure A7 – Major City Borders Are Stable Over Time, Suggesting Little Risk of Measurement Error from Changing Longitude / Latitude Centroids




*Notes:* The purpose of this figure is to provide evidence that borders for major cities are not changing substantially over time. Given that the longitude and latitude of each birth place is assigned based on city/place centroids, substantially changing borders over time could introduce time-varying measurement error via centroids that are inaccurate for certain birth cohorts. However, this concern is mitigated by the fact that these city borders—presented for all California cities with available GIS data dating back to 1930—are highly similar. (Note that 1980 was chosen as a comparison year as it is the closest year to the *end* of our sample period for which there was place shapefile data). Instead of changing borders of principal cities, as these major metropolitan areas grew, they simply created new cities and municipalities with their own associated centroids. Accordingly, time-varying measurement error due to changing centroids is likely not an issue.

*Source:* Author calculations using shapefile data obtained from the Urban Transition Historical GIS Project (Logan and Zhang, 2020) and IPUMS NHGIS (Manson et al., 2021).

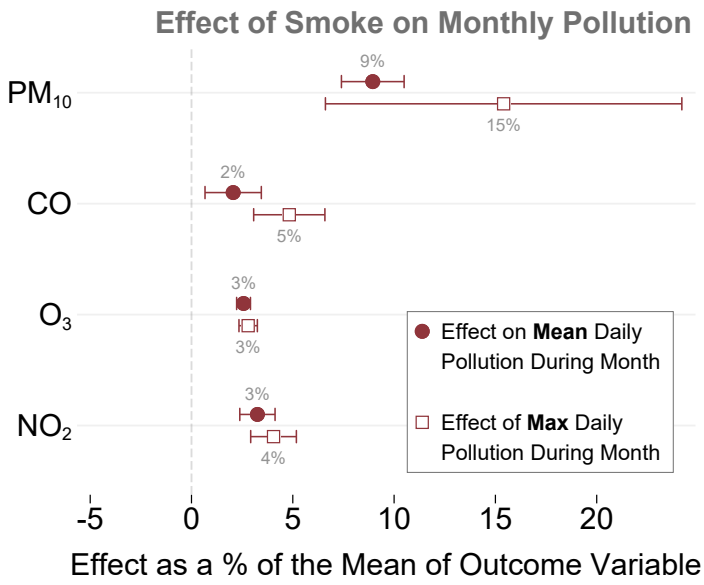
Figure A8 – Descriptive Statistics: Comparison of Adjusted and Unadjusted Death Rates per Numident to Expected Cohort Death Rates




*Notes:* The purpose of this figure is to contrast how unadjusted and adjusted death rates per the Numident file compare with death rates from SSA Cohort Life Tables. As displayed in the figure, adjusted death rates are much closer to expected values and are more consistent over different cohorts. All values less than 0.5 are bottom coded for disclosure avoidance purposes. See Section 3.3 and Appendix Section A for further discussion.

*Source:* Author calculations using Restricted Census Numident file and SSA Cohort Life Tables.

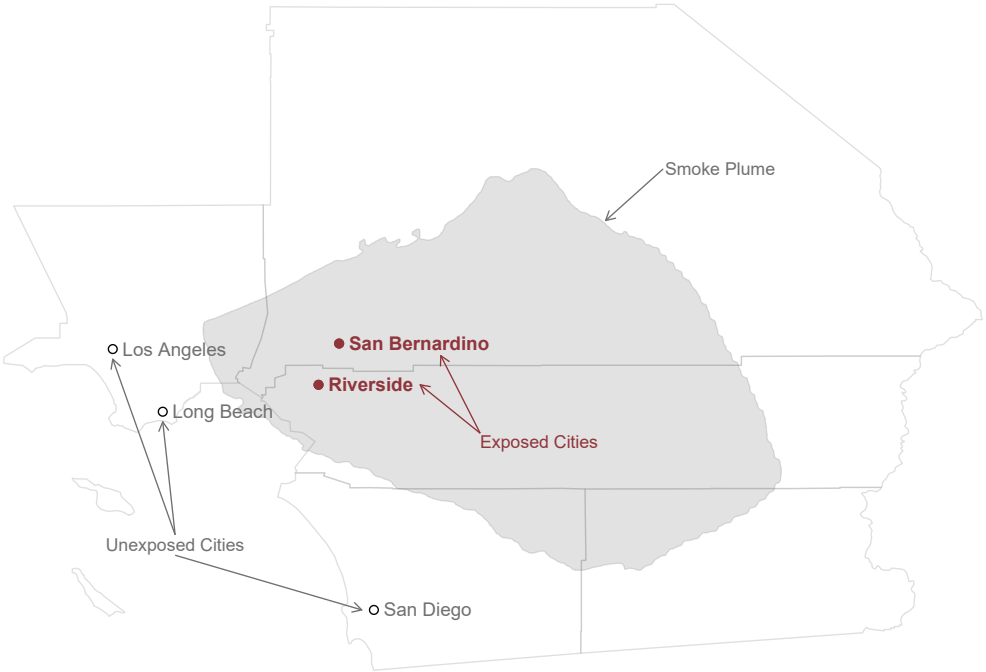
Figure A9 – Validation of Treatment: Modeled Wildfire Smoke Exposure Increases Monthly Measures of Various Pollutants




*Notes:* The purpose of this figure is to display the impact of modeled smoke plume coverage on monthly pollution levels across a range of pollutants (described on the $y$-axis). Each point represents the $\beta_0$ coefficient from the regression detailed by Equation 4 which has been scaled by the mean of the given pollutant for comparability across different pollution types. Each scaled estimate is presented with corresponding 95% confident intervals.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), and Daily Summary Data Files from the EPA Data Mart.

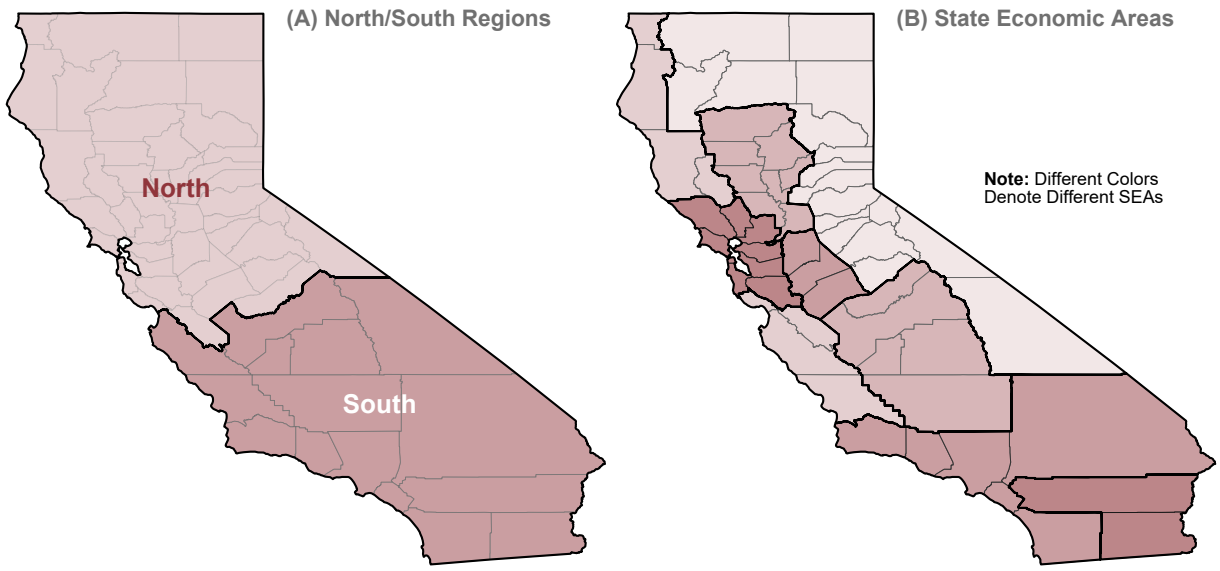
Figure A10 – Illustration of Treatment Definition




*Notes:* The purpose of this figure is to illustrate how treatment is assigned. A hypothetical/modeled smoke plume is presented in the figure. The cities that are considered to be treated—those that spatially intersect with the smoke plume—are colored maroon, while the unexposed cities—those that are outside of the smoke plume—are colored in black/gray.

*Source:* Author simulations using HYSPLIT model output (Stein et al., 2015).

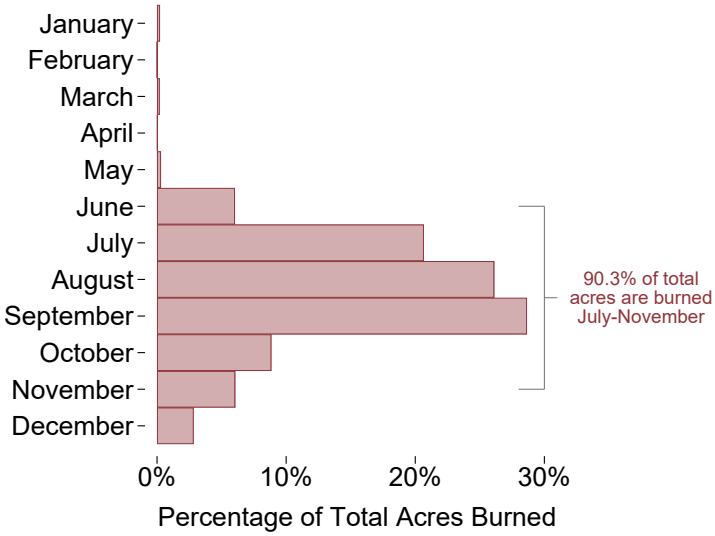
Figure A11 – Illustration of Regional Definitions




*Notes:* The purpose of this figure is to display different regional definitions used in our analysis. As discussed in Section 5 our core specification utilizes year-of-birth by region (Northern and Southern California) fixed effects. The counties included in Northern and Southern California are included in Panel A. In our robustness checks (Section 8), we instead utilize year-of-birth by State-Economic-Area fixed effects. The counties included in the different State Economic Areas are included in Panel B.

*Source:* Author calculations using data from the U.S. Census Bureau and Ruggles et al. (2021).

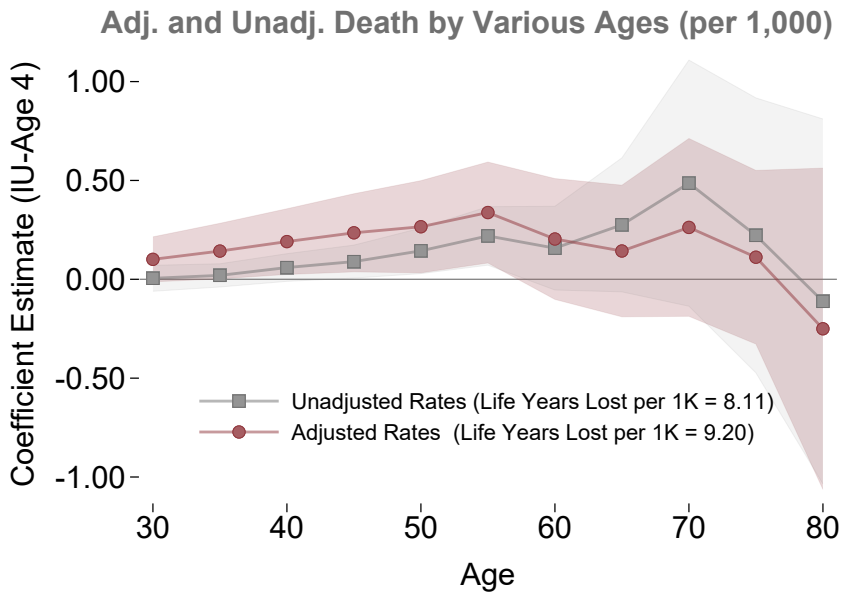
Figure A12 – Descriptive Statistics: Fire Seasonality




*Notes:* The purpose of this figure is to display seasonality in wildfire acres burned.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018).

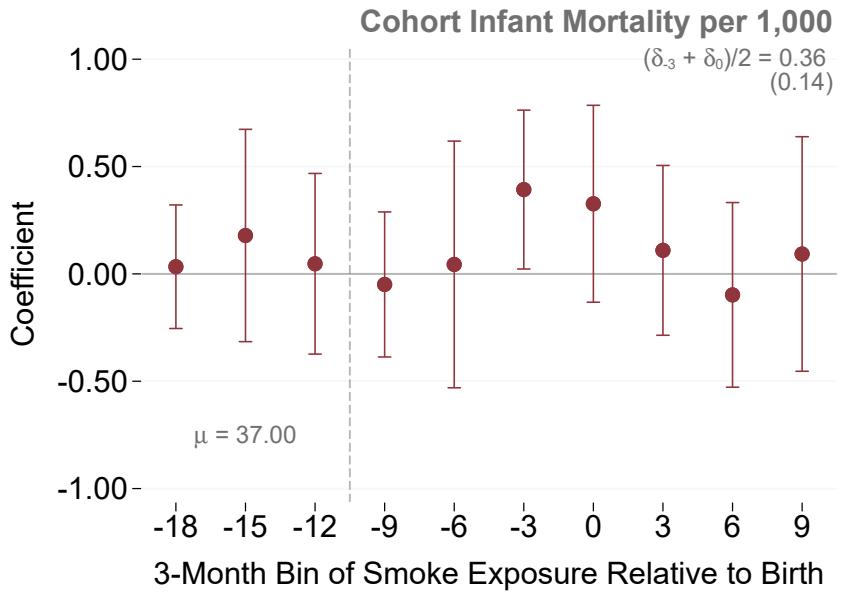
Figure A13 – Robustness: Unadjusted and Adjusted Mortality Measures Display Similar Effects




*Notes:* The purpose of this figure is to show similarities in death-by-age results when using unadjusted and adjusted measures. See Figure 4 and accompanying text for further information on figure construction and discussion of results.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), and Restricted Census Numident.

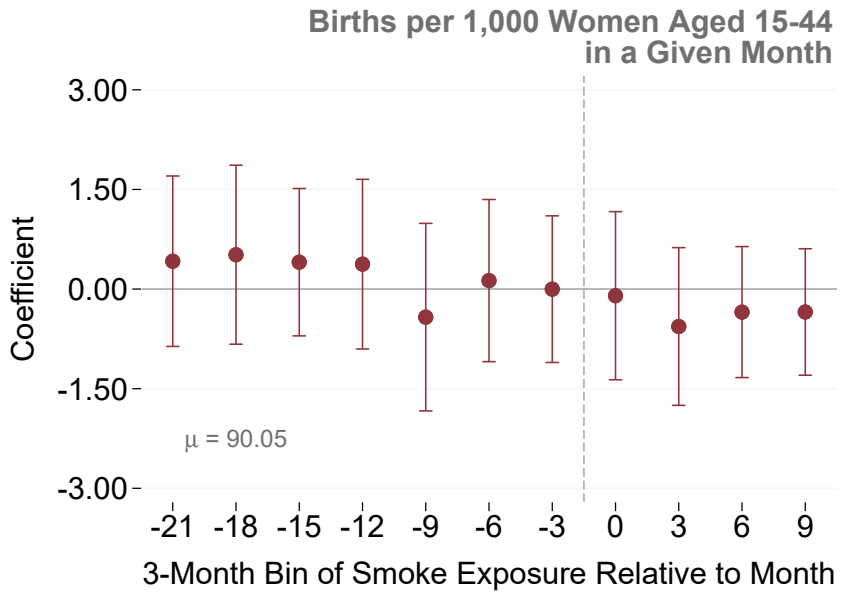
Figure A14 – Additional Results: Wildfire Smoke Exposure Increases Infant Mortality




*Notes:* The purpose of this figure is to assess the impact of wildfire smoke on cohort infant mortality. Within the figure, each coefficient displays the effect of an additional smoke month within three-month bins, along with 95% confidence intervals. In the upper right-hand corner, the average effect during the critical period—considered to be the three months before ($\delta_{-3}$) and after ($\delta_0$) birth—is presented alongside its standard error. The mean is presented in the lower left-hand corner for reference.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), California Department of Health Vital Statistics, and Natality/Mortality data from Bailey et al. (2018).

Figure A15 – Additional Results: There is No Detectable Effect of Wildfire Smoke Exposure on Fertility Timing or Overall Fertility



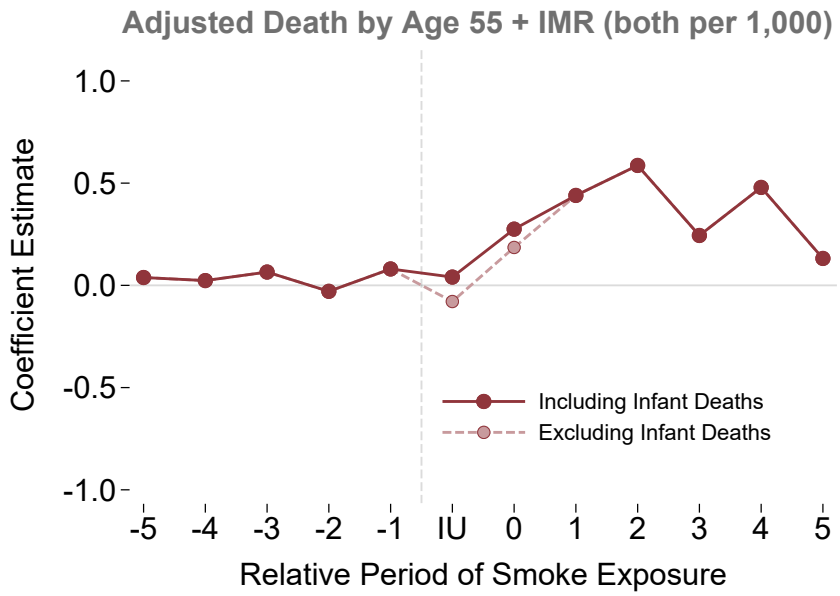

*Notes:* The purpose of this figure is to estimate the effect of wildfire smoke exposure on fertility, as measured by births per 1,000 women aged 15-44. To do so, we estimate the following equation:

$$BirthsPer1000_{Cb} = \mu_C + \lambda_{r(C),b} + \sum_{j=-21,-18,\ldots,9} \delta_j Smoke_{C,b+j} + \varepsilon_{Cb},$$

which is very similar to the equation used to estimate infant mortality in Appendix Section B.1 (see text of that section for more details). Within the figure, coefficients from this equation are plotted with accompanying 95% confidence intervals. The mean of the outcome variable is presented in the bottom-left corner for reference. As displayed by the figure, there do not appear to be any statistically-detectable effects on fertility timing or overall fertility.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), California Department of Health Vital Statistics, and Natality/Mortality data from Bailey et al. (2018).

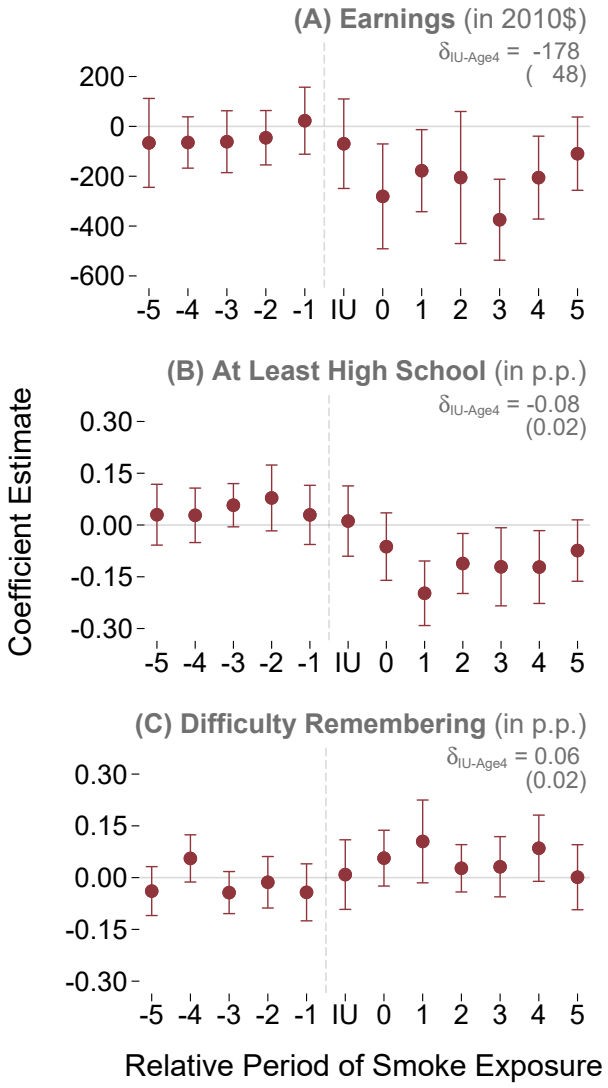
Figure A16 – Additional Results: Reductions in Longevity with Infant Mortality Added




*Notes:* The purpose of this figure is to display the coefficients from Figure 4 Panel A with infant mortality results (discussed in Appendix Section B.1) added in. Confidence intervals are omitted from this figure for clarity of presentation.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, California Department of Health Vital Statistics, and Natality/Mortality data from Bailey et al. (2018).

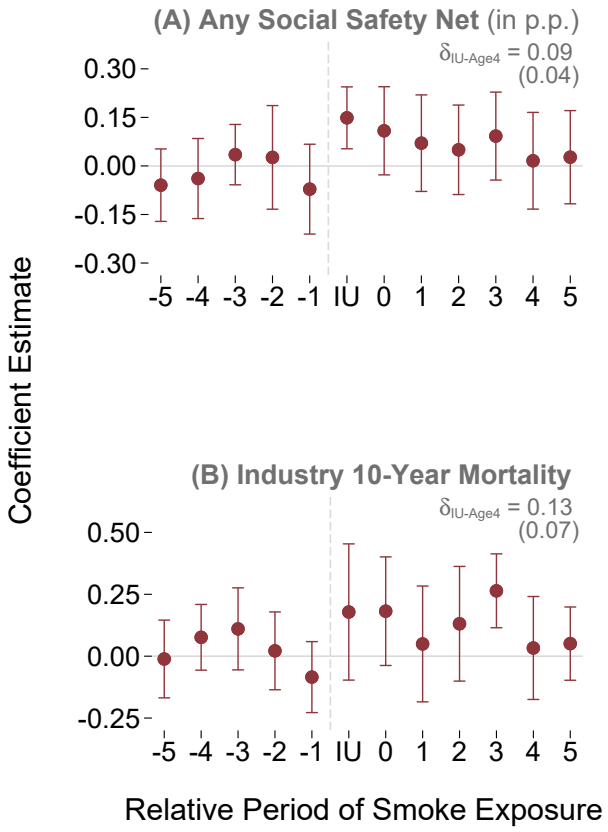
Figure A17 – Additional Results: Early-Life Wildfire Smoke Exposure Reduces Earnings and High School Completion; Increases Cognitive Difficulty




*Notes:* The purpose of this figure is to display the impact of childhood air pollution exposure from wildfires on a selected group of long-run outcomes. See notes to Figure 4 for more detail on graph construction and Section 7 for further discussion.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, and Restricted American Community Surveys.

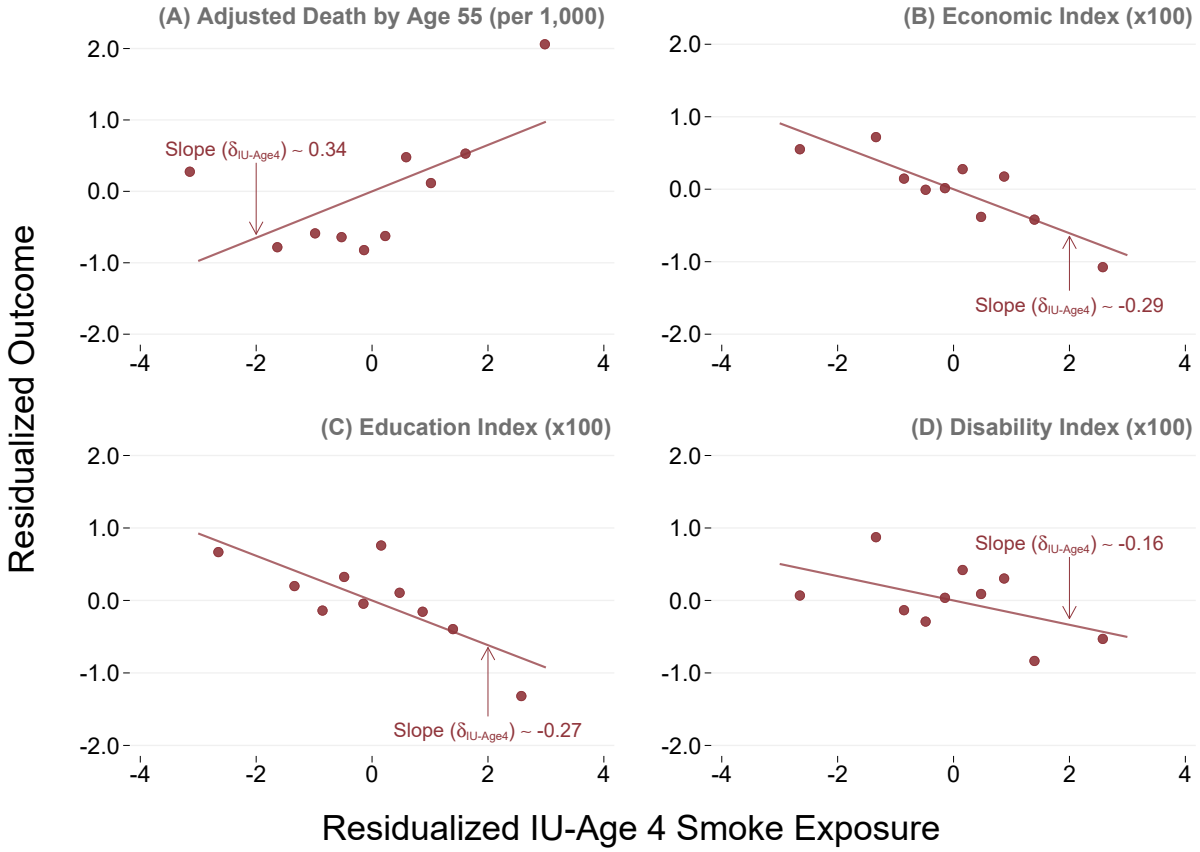
Figure A18 – Additional Results: Early-Life Wildfire Smoke Exposure Increases Use of Social Safety Net; Induces Sorting into Higher-Mortality Industries




*Notes:* The purpose of this figure is to display the impact of childhood air pollution exposure from wildfires on a selected group of long-run outcomes. See notes to Figure 4 for more detail on graph construction and Sections 7 and 9 for further discussion.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, Restricted American Community Surveys, and Restricted Full-count Decennial Censuses.

Figure A19 – Robustness: Residual Plots Demonstrate that Impacts of Wildfire Exposure are Generally Linear




*Notes:* The purpose of this figure is to show how each of our main outcomes varies with our treatment. To do so, we residualize our baseline fixed effect controls from both the outcome and our summary treatment measure as follows:

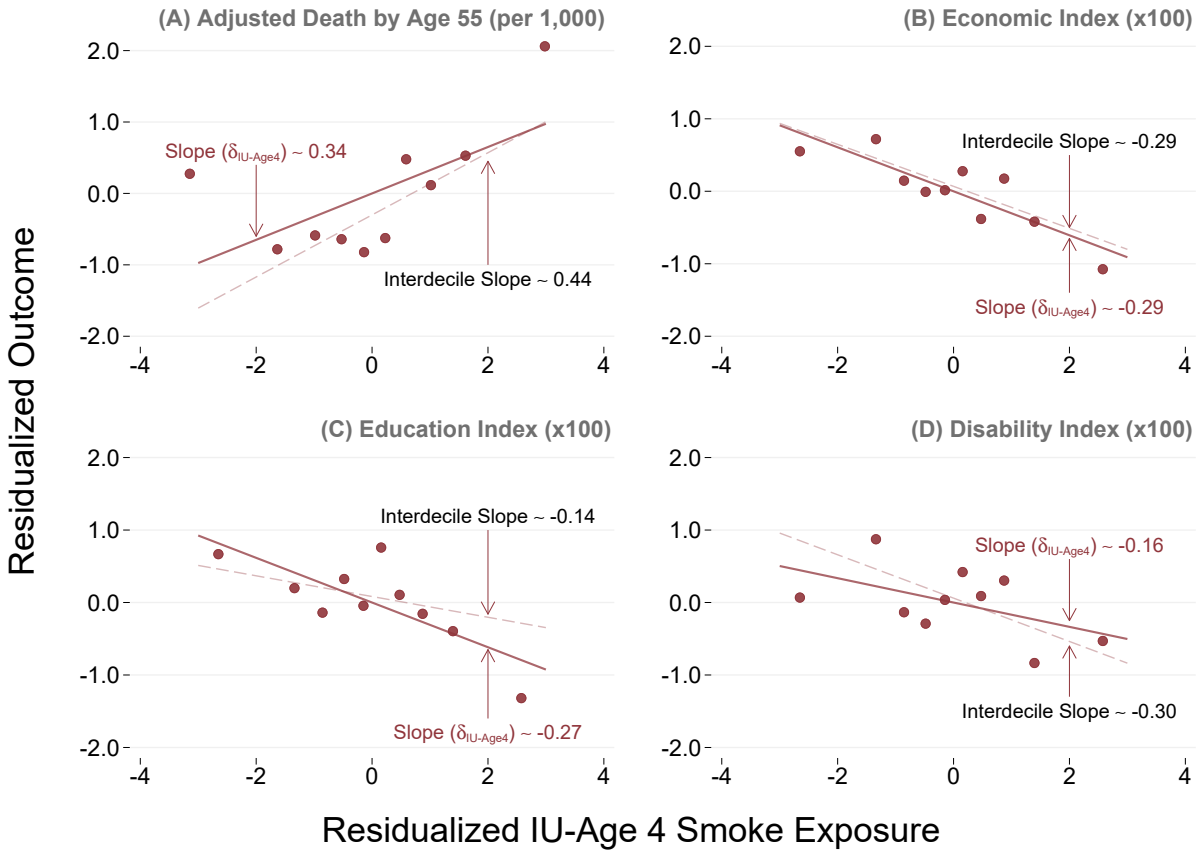
$$y_{cba} = \mu_c + \lambda_{y(b),r(c)} + \alpha_{m(b)} + \psi_{y(b),a} + \epsilon_{cba}$$

$$Smoke_{cb}^{\text{IU-Age4}} = \mu_c + \lambda_{y(b),r(c)} + \alpha_{m(b)} + \psi_{y(b),a} + \xi_{cba}.$$

We then collapse the residualized summary treatment ($\tilde{\xi}$) into deciles and plot it against the residualized outcome ($\tilde{\varepsilon}$). The fitted line between $\tilde{\xi}$ and $\tilde{\varepsilon}$ gives the coefficient for our treatment. The general shape of the points within the figure supports our choice of functional form—the dots generally follow a linear pattern, underscoring that the dose-response for our treatment is linear in months of smoke exposure. The only meaningful deviations from this trend occur at the very tails of the distribution. However, as illustrated by Appendix Figure A20, removal of these points does not materially change the estimates.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, and Restricted American Community Surveys.

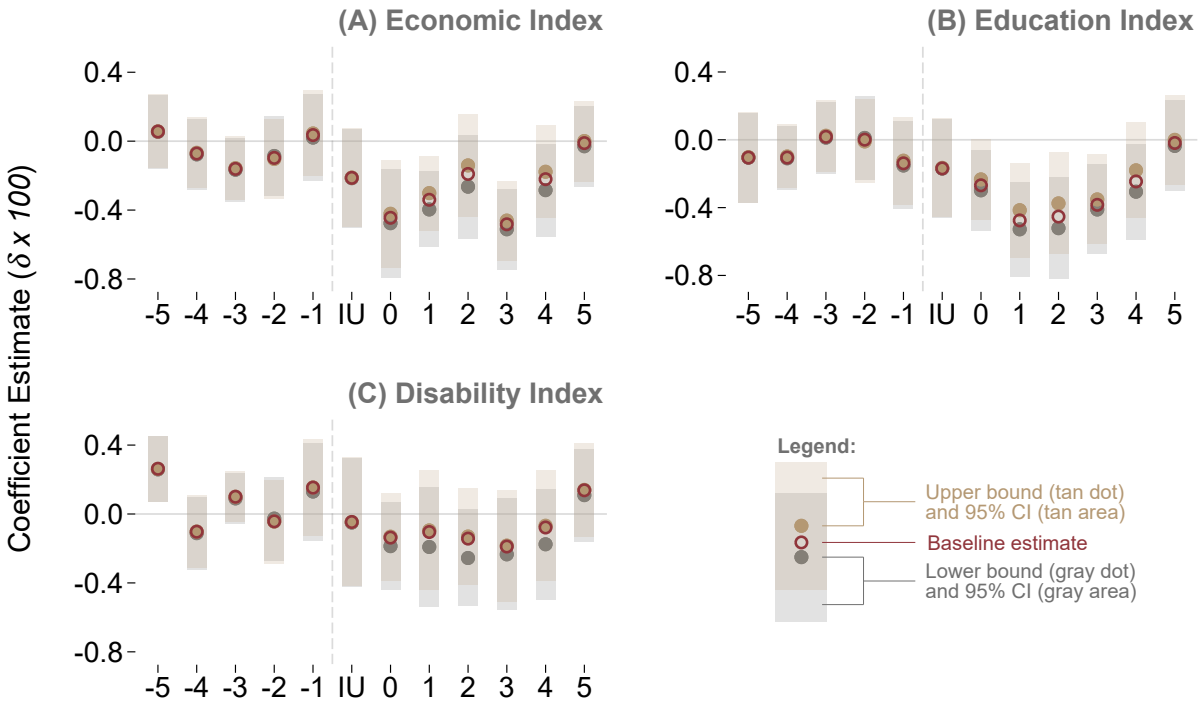
Figure A20 – Robustness: Residual Plots Demonstrate that Impacts of Wildfire Exposure are Not Driven by the Tails of the Distribution




*Notes:* The purpose of this figure is to further support the analysis displayed in Appendix Figure A19 and discussed in the accompanying notes. Specifically, we modify the figure to fit lines and calculate slopes using only the points in the 2nd through 9th deciles (the "interdecile slope") to demonstrate that our estimates are not driven by extreme points in the distribution. See Appendix Figure A19 for more detail, including information regarding the construction of the graphs.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, and Restricted American Community Surveys.

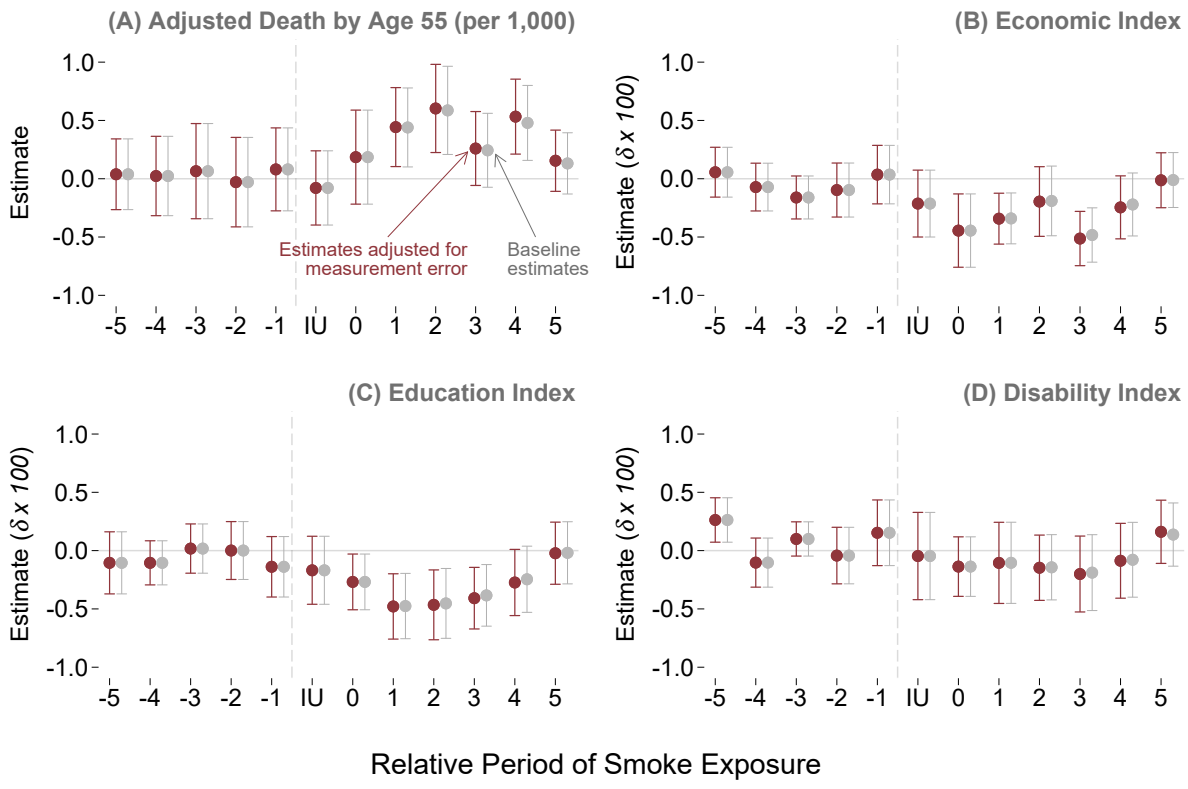
Figure A21 – Robustness: Adjusting for Mortality Selection Does Not Qualitatively Affect Estimates




*Notes:* The purpose of this figure is to display the bounded estimates for mortality selection discussed in Section 8 and Appendix Section B.3. Within the figure, upper- and lower-bounds of the potential impact of mortality selection are presented, alongside baseline estimates (i.e., those from Figure 5). See the text of the referenced sections for further description.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, Restricted American Community Surveys, California Department of Health Vital Statistics, and Natality/Mortality data from Bailey et al. (2018).

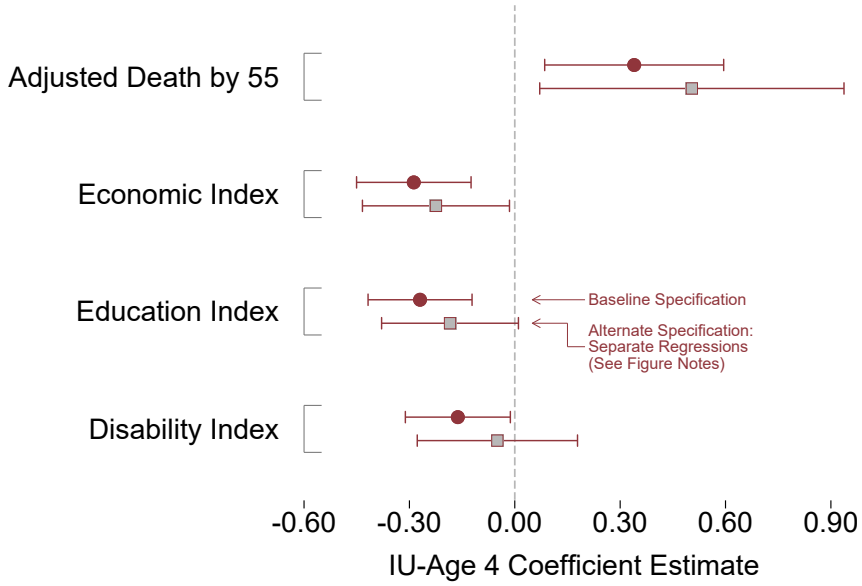
Figure A22 – Robustness: Adjusting for Migration-Specific Measurement Error Does Not Qualitatively Affect Estimates




*Notes:* The purpose of this figure is to compare our original estimates with those that have been adjusted for migration-specific measurement error. Within the figure, the gray estimates display the baseline estimates whereas the maroon estimates display the estimates when they have been adjusted for measurement error using the techniques described in Section 8 and Appendix Section B.4. See the text of those sections for further detail.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, and Restricted American Community Surveys, 1940 Census 100% Sample (Ruggles et al., 2021).

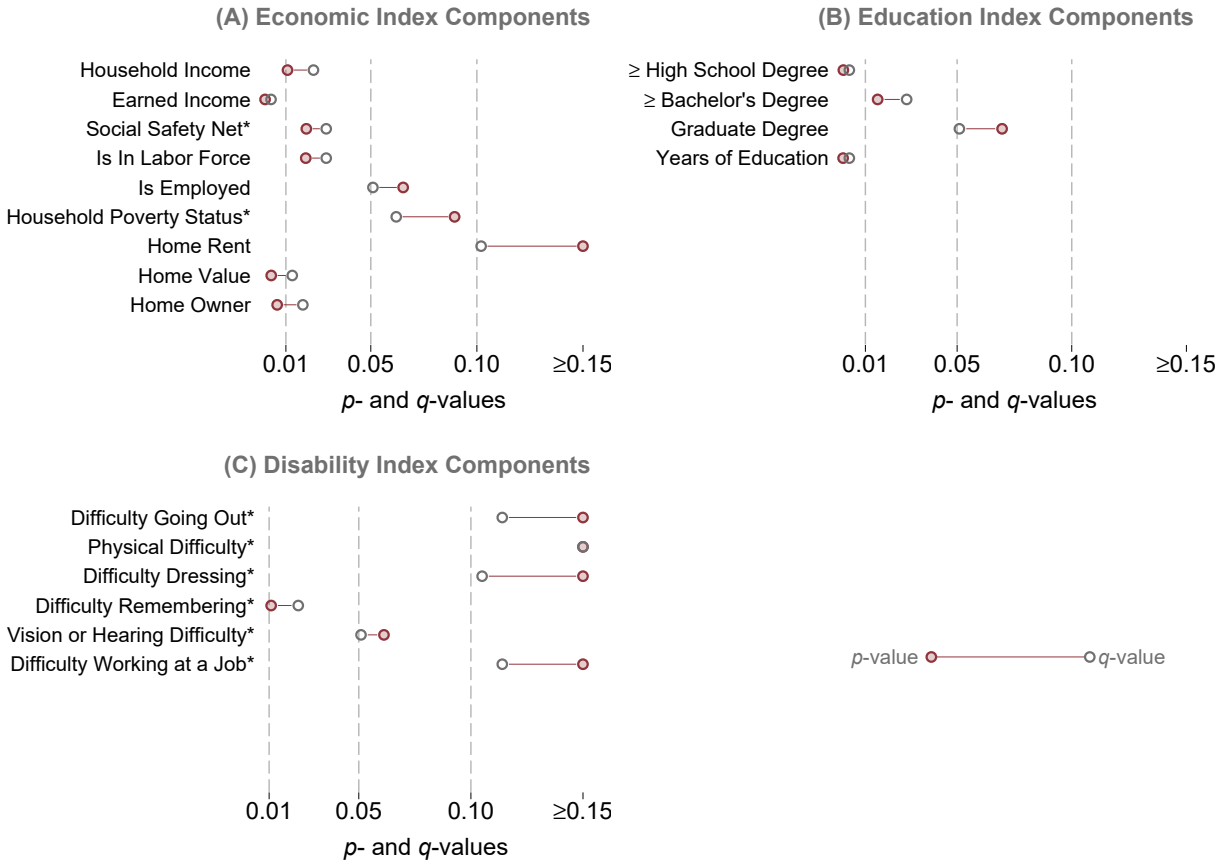
Figure A23 – Robustness: Using Alternative Estimation Strategies Do Not Result in Materially Different Estimates




*Notes:* The purpose of this figure is to display our estimates from the alternative estimation strategy discussed in Section 8 as compared to our original estimates. This alternative estimation strategy involves separately estimating wildfire smoke effects for each year-of-birth cohort and combining these estimates via simple average. Standard errors are calculated using 100 cluster bootstrap iterations. As further discussed in the text, estimates are qualitatively similar, although the alternative estimates are substantially less precise.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, and Restricted American Community Surveys.

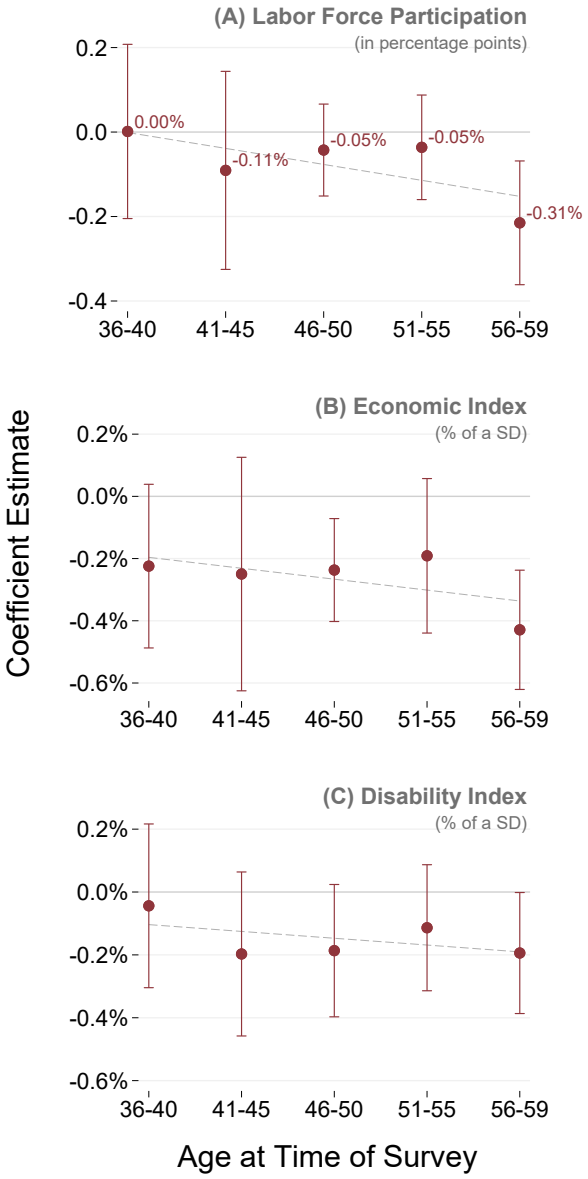
Figure A24 – Robustness: Multiple Hypothesis Testing Does Qualitatively Affect Conclusions About Statistical Significance for Index Components




*Notes:* The purpose of this figure is to contrast the $p$-values for our estimates with sharpened $q$-values (Benjamini and Yekutieli, 2001; Benjamini et al., 2006; Anderson, 2008) that adjust for multiple hypothesis testing. As noted by Anderson (2008), the $q$-values can be *smaller* than unadjusted $p$-values when there are many true rejections of the null hypothesis. In this case, then the method can "tolerate" some false rejections as well. Note that this tends to occur for estimates with large $p$-values, such as those in Panel C, which do not achieve traditional levels of statistical significance even after adjustment.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, and Restricted American Community Surveys.

Figure A25 – Supplemental Results: The Effects of Early-Life Wildfires Smoke Data on Labor Force Participation, Economic Index, and Disability Index are Increasing Proportionally with Age



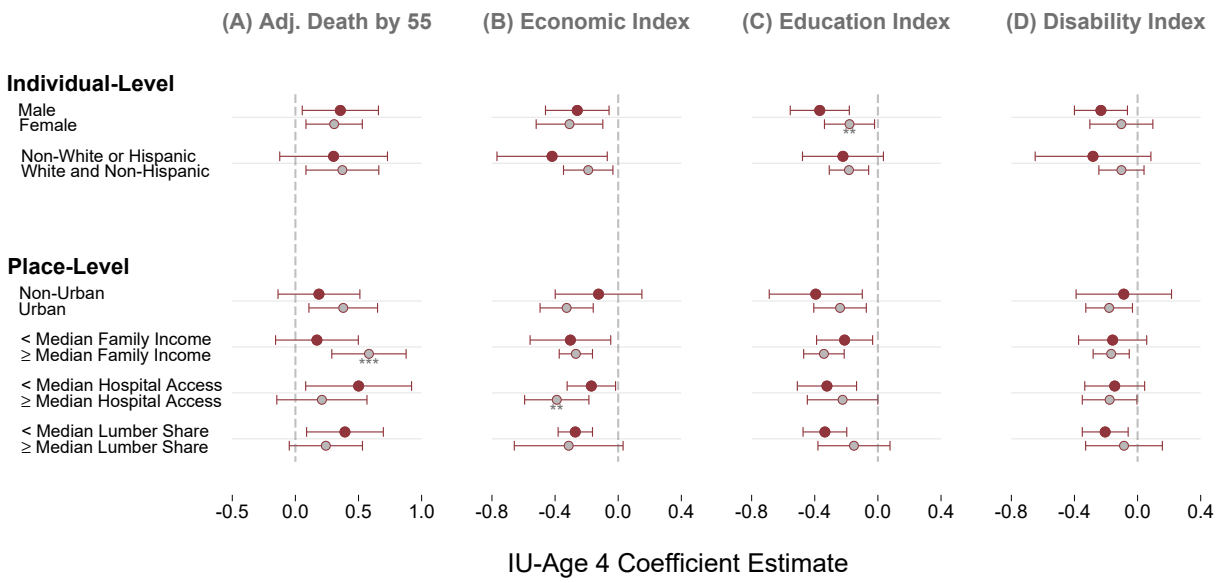

*Notes:* The purpose of this figure is to display how the effects of childhood wildfire smoke exposure evolve as individuals age. Within the figure, each point represents a separate estimate of $\delta_j$ with associated 95% confidence intervals, from the following equation:

$$Y_{cba} = \mu_c + \lambda_{y(b),r(c)} + \alpha_{m(b)} + \psi_{y(b),a} + \sum_{j \in \mathcal{J}} \delta_j \cdot (AgeBin_j \times Smoke_c^{\text{IU-Age 4}}) + \varepsilon_{cba},$$

For further context, a line of best fit for the coefficients, along the effect as a percent of the age-specific mean is also presented. (Note that the percent of the age-specific mean is *not* presented for the indexes, as they are created to have a mean of zero in the aggregate.)

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, and Restricted American Community Surveys.

Figure A26 – Supplemental Results: There Are Few Statistically Significant Differences in the Effect of Early-Life Wildfire Smoke Across Groups

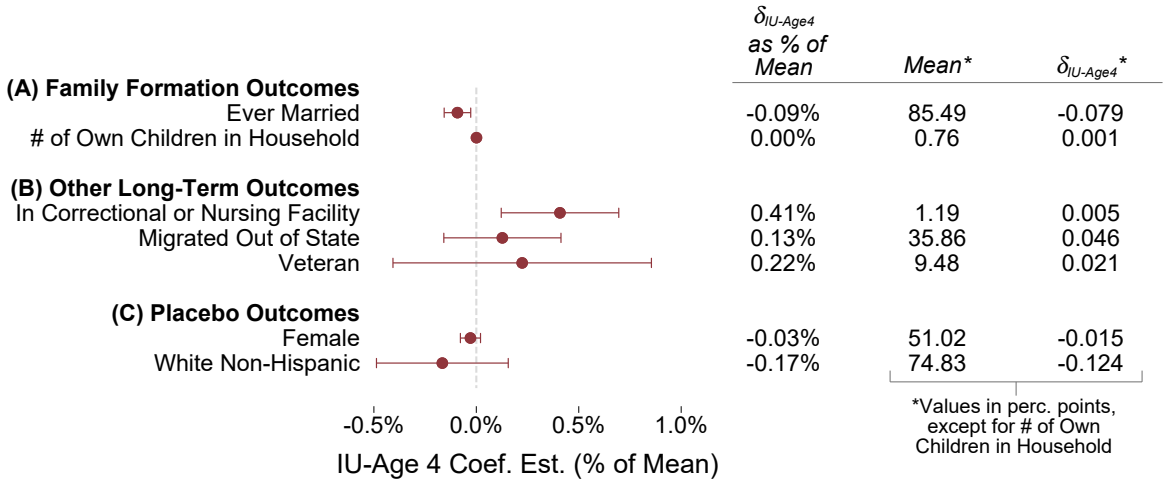



*Notes:* The purpose of this figure is to display the results of our heterogeneity analysis across groups. Within the figure, there are two sets of estimates presented for each regression detailed by Equation B2: the baseline effect ($\delta_0$) and the *combined* effect ($\delta_0 + \delta_1$) which captures the total effect on the interaction category. Both the baseline effect and combined effect are presented with their 95% confidence intervals. Additionally, estimate pairs that are statistically distinct from each other—i.e., where the interaction term, $\delta_1$, achieves statistical significance—are denoted as follows:

* Significant at 90% level
** Significant at 95% level
*** Significant at 99% level

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, Restricted American Community Surveys, 1930 Census 100% Sample (Ruggles et al., 2021), and American Hospital Association Data (Finkelstein, 2007).

Figure A27 – Supplemental Results: Early-Life Wildfire Smoke Exposure Does Not Significantly Affect Family Formation, Migration, or Placebo Outcomes, but Does Increase Institutionalization




*Notes:* The purpose of this figure is to display the impact of childhood air pollution exposure from wildfires on outcomes that do not fall into the index categories discussed in Section 7. Within the figure, estimates—scaled to represent the effect as a percent of the mean—are presented alongside 95% confidence intervals. The tabular effect as a percent of the mean is presented to the first column on the right-hand size of the figure, along with the mean and level (non-scaled) estimate in the second and third columns, respectively.

The outcomes considered fall into three groups. First, we consider family formation outcomes, i.e., whether an individual was ever married (source: ACS) and the number of own children in the household at the time of survey (source: 2000 and 2010 Decennial Censuses). The outcomes in this group are either statistically or economically insignificant. The next group includes outcomes includes whether an individual lives out of state in adulthood (source: 2000 and 2010 Decennial Censuses), their veteran status (source: ACS), or whether they reside in institutional group quarters, the majority of which are correctional or nursing facilities (source: 2010 Decennial Census only; this variable is not available in the Restricted 2000 Decennial Census). Note that correctional and nursing facilities are grouped together because the Census Bureau did not permit RDC researchers to disclose results using sub-categories of the institutional group quarters variable during the time period in which these results were disclosed. The final group includes placebo outcomes—i.e., outcomes which should not be affected by early-childhood wildfire smoke. We do not find any effect on these outcomes, as expected.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, Restricted American Community Surveys, and Restricted Full-count Decennial Censuses.

Table A1 – Summary Statistics

| Variable | Mean | Standard Deviation |
|---|---|---|
| *Economic Outcomes* (*American Community Survey, Ages 35-59*) | | |
| Household Income | $ 99,367 | $ 87,043 |
| Earned Income* | $ 44,087 | $ 54,488 |
| Social Safety Net** | 14% | 34% |
| Is In Labor Force | 76% | 43% |
| Is Employed | 71% | 45% |
| Household Poverty Status | 11% | 31% |
| Home Rent | $ 1,176 | $ 614 |
| Home Value | $ 418,807 | $ 340,128 |
| Home Owner | 73% | 45% |
| *Education Outcomes* (*American Community Survey, Ages 35-59*) | | |
| At Least High School Degree | 94% | 24% |
| At Least Bachelor's Degree | 29% | 45% |
| Graduate Degree | 10% | 30% |
| Years of Education | 14.80 | 2.62 |
| *Disability Outcomes* (*American Community Survey, Ages 35-59*) | | |
| No Disabilities*** | 84% | 37% |
| Difficulty Going Out (Doing Errands Alone) | 5% | 23% |
| Physical Difficulty (Walking or Climbing Stairs) | 10% | 30% |
| Difficulty Dressing or Bathing | 3% | 18% |
| Difficulty Remembering, Concentration or Making Decisions | 6% | 24% |
| Vision or Hearing Difficulty | 5% | 21% |
| Difficulty Working at a Job | 11% | 31% |
| Number of Disabilities*** | 0.32 | 0.85 |
| *Mortality per 1,000* (*Numident*) | | |
| Adjusted Deaths by Age 35 | 27.22 | |
| Adjusted Deaths by Age 45 | 50.19 | |
| Adjusted Deaths by Age 55 | 94.96 | |
| Adjusted Deaths by Age 65 | 176.27 | |
| Adjusted Deaths by Age 75 | 321.90 | |

*Notes:* The purpose of this table is to detail means and standard deviations for key outcome variables included in our analysis.

* Variable is equal to wages + self-employment income.
** Variable is equal to 1 if an individual has any utilization of SNAP, Medicaid, SSI, or public assistance.
*** These variables are not included in the disability index and are only provided for context.

*Source:* Public-Use American Community Surveys (Ruggles et al., 2021) and SSA Numident.

Table A2 – Main Results: Table of Estimates

| Outcome | Estimate | Standard Error | *t*-Statistic |
|---|---|---|---|
| *Panel A: Mortality Outcomes (per 1,000)* | | | |
| Adjusted Death by Age 30 | 0.10 | (0.06) | 1.73 |
| Adjusted Death by Age 35 | 0.14 | (0.07) | 1.99 |
| Adjusted Death by Age 40 | 0.19 | (0.08) | 2.25 |
| Adjusted Death by Age 45 | 0.24 | (0.10) | 2.33 |
| Adjusted Death by Age 50 | 0.27 | (0.12) | 2.23 |
| Adjusted Death by Age 55 | 0.34 | (0.13) | 2.59 |
| Adjusted Death by Age 60 | 0.20 | (0.16) | 1.31 |
| Adjusted Death by Age 65 | 0.14 | (0.17) | 0.84 |
| Adjusted Death by Age 70 | 0.26 | (0.23) | 1.14 |
| Adjusted Death by Age 75 | 0.11 | (0.22) | 0.50 |
| Adjusted Death by Age 80 | -0.25 | (0.41) | -0.60 |
| *Panel B: Index Outcomes ($\delta \times 100$)* | | | |
| Economic Index | -0.29 | (0.08) | -3.46 |
| Household Income | -0.34 | (0.13) | -2.55 |
| Earned Income | -0.38 | (0.10) | -3.73 |
| Social Safety Net* | -0.33 | (0.14) | -2.33 |
| Is In Labor Force | -0.22 | (0.09) | -2.34 |
| Is Employed | -0.19 | (0.10) | -1.84 |
| Household Poverty Status* | -0.23 | (0.14) | -1.70 |
| Home Rent | -0.38 | (0.27) | -1.43 |
| Home Value | -0.31 | (0.10) | -2.96 |
| Home Owner | -0.26 | (0.09) | -2.75 |
| Education Index | -0.27 | (0.08) | -3.57 |
| At Least High School Degree | -0.32 | (0.09) | -3.54 |
| At Least Bachelor's Degree | -0.22 | (0.09) | -2.42 |
| Graduate Degree | -0.19 | (0.10) | -1.81 |
| Years of Education | -0.35 | (0.10) | -3.63 |
| Disability Index* | -0.16 | (0.08) | -2.13 |
| Difficulty Going Out (Doing Errands Alone)* | -0.12 | (0.10) | -1.30 |
| Physical Difficulty (Walking or Climbing Stairs)* | -0.08 | (0.13) | -0.65 |
| Difficulty Dressing or Bathing* | -0.12 | (0.09) | -1.38 |
| Difficulty with Remembering/Concentrating/Decisions* | -0.30 | (0.12) | -2.55 |
| Vision or Hearing Difficulty* | -0.26 | (0.14) | -1.87 |
| Difficulty Working at a Job* | 0.16 | (0.15) | 1.11 |

*Notes:* The purpose of this table is to detail coefficient estimates presented in Figures 4 and 5, along with standard errors and *t*-statistics.
* Indicates that variable has been reverse coded so that higher coefficient values correspond with "better" outcomes.

*Source:* Author calculations using the Fire Perimeter Database (California Department of Forestry and Fire Protection, 2018), HYSPLIT model output (Stein et al., 2015), Restricted Census Numident, Restricted 2000 Decennial Census 1-in-6 Sample, and Restricted American Community Surveys.

Table A3 – Context for Main Results: Level Estimates

| Outcome | Estimate | Mean | Effect as % of Mean | Unit | Note |
|---|---|---|---|---|---|
| *Economic Outcomes* | | | | | |
| Household Income | -$280 | $99,367 | -0.28% | Dollars | * |
| Earned Income | -$178 | $44,087 | -0.40% | Dollars | * |
| Social Safety Net | 0.09 | 13.62 | 0.66% | Percentage Points | * |
| Is In Labor Force | -0.10 | 76.01 | -0.13% | Percentage Points | * |
| Is Employed | -0.09 | 71.40 | -0.12% | Percentage Points | |
| Household Poverty Status | 0.07 | 10.73 | 0.66% | Percentage Points | |
| Home Rent | -$2 | $1,176 | -0.20% | Dollars | |
| Home Value | -$1,044 | $418,807 | -0.25% | Dollars | |
| Home Owner | -0.12 | 72.78 | -0.16% | Percentage Points | |
| *Education Outcomes* | | | | | |
| At Least High School Degree | -0.08 | 93.94 | -0.09% | Percentage Points | * |
| At Least Bachelor's Degree | -0.10 | 28.57 | -0.34% | Percentage Points | |
| Graduate Degree | -0.06 | 10.35 | -0.55% | Percentage Points | |
| Years of Education | -0.01 | 14.80 | -0.06% | Years | |
| *Disability Outcomes* | | | | | |
| Difficulty Going Out | 0.03 | 5.48 | 0.51% | Percentage Points | |
| Physical Difficulty | 0.02 | 9.65 | 0.25% | Percentage Points | |
| Difficulty Dressing | 0.02 | 3.24 | 0.68% | Percentage Points | |
| Difficulty Remembering | 0.06 | 6.23 | 1.01% | Percentage Points | * |
| Vision or Hearing Difficulty | 0.06 | 4.70 | 1.17% | Percentage Points | |
| Difficulty Working at a Job | -0.05 | 10.81 | -0.47% | Percentage Points | |

*Notes:* The purpose of this table is to provide level coefficients for the components of the indices displayed in Figure 5 (which displays coefficients that are scaled to be in terms of a percentage of the outcome's standard deviation). Within this table, most level estimates are obtained by multiplying the coefficient estimate by the applicable standard deviation. However, outcomes denoted with a "*" are outcomes for which the level effect has been separately estimated and disclosed.

*Source:* Author calculations using the Restricted and Public-Use (Ruggles et al., 2021) American Community Surveys.

Table A4 – Supplemental Results: Wildfire Smoke Does Not Have Detectable Effects on Contemporaneous Local Economic Conditions or Net Migration

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | **% in Labor Force** | **% Employed** | **% Home Owners** | **Home Value** | **Family Income** | **Net Migration** |
| $Smoke^{t',t}$ | 0.00 | 0.02 | -0.04 | -$125 | -$35 | -0.07 |
| | (0.04) | (0.04) | (0.06) | (106) | (51) | (0.63) |
| 95% CI: Upper Bound | 0.09 | 0.09 | 0.09 | $83 | $65 | 1.16 |
| 95% CI: Lower Bound | -0.08 | -0.06 | -0.16 | -$332 | -$135 | -1.31 |
| Outcome Mean | 60.60 | 55.90 | 54.40 | $93,056 | $43,619 | 24.30 |
| Observations | 232 | 232 | 172 | 170 | 169 | 232 |

*Notes:* The purpose of this table is to demonstrate the impact of an additional smoke month on a variety of outcomes that are *contemporaneous* to the smoke exposure. Each column in the figure represents a different outcome variable estimated from Equation B1, with associated standard errors clustered at the county level. Note that columns 1-3 and 6 are in percentage points, while columns 4 and 5 are denominated in 2010 dollars. See the text of Appendix Section B.5 for more detail.

*Source:* Author calculations using the 1930-1970 Decennial Census (Ruggles et al., 2021), Census County Book Data (U.S. Census Bureau, 2012), and 1934 IRS Statistics on Income.