Cliometrica
https://doi.org/10.1007/s11698-024-00297-0

ORIGINAL PAPER

The impact of wildfire smoke exposure on excess mortality and later-life socioeconomic outcomes: the Great Fire of 1910

Sarah Meier[1] · Eric Strobl[2,3] · Robert J. R. Elliott[3]

Received: 10 July 2024 / Accepted: 31 August 2024

Abstract

The Great Fire of 1910 in the northwestern United States burnt more than 1.2 million hectares in just two days and stands as one of the largest wildfires ever recorded. While it is known for having led to the introduction of a rigorous fire suppression regime that lasted for much of the twentieth century, it also generated a considerable amount of smoke far beyond the burnt areas that is likely to have impacted the health of those exposed. This paper examines the short- and long-term impact of this fire-sourced smoke pollution on children, combining historical data with smoke emission and dispersion modelling. The econometric results indicate a 119% increase in excess mortality during the week of the fire and a decrease of 4–14% in later-life socioeconomic status scores 20 and 30 years after the event. This research offers novel insights into wildfire smoke repercussions on health and long-run human capital formation in a setting where avoidance behaviour was minimal.

Keywords Wildfire · Air pollution · Health effects · Human capital · Socioeconomic outcomes

JEL Classification I1 · N3 · N5 · Q5

This research has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement Innovative Training Networks No. 860787. In addition, Sarah Meier gratefully acknowledges the support of the Dragon Capital Chair on Biodiversity Economics funding her current Postdoctoral Research Fellowship. We would additionally like to thank Ethan Addicott, David Maddison, Sefi Roth, Randall Walsh, as well as seminar and conference participants at the Nova School of Business and Economics, the Vrije Universiteit Amsterdam, the University of Birmingham, the University of Exeter, the LSE Grantham Institute, AERE, UEA Copenhagen, LAGV Marseille, and EAERE for their insightful comments and suggestions at various stages of the manuscript.

Extended author information available on the last page of the article

Published online: 11 November 2024

1 Introduction

Large environmental shocks are often remembered for being defining events in history, causing institutional, political, cultural, social, and economic change (Ljungqvist et al. 2021; Harper 2021; Degroot et al. 2022; Frankopan 2023).[1] The Great Fire of 1910 in the United States (US) certainly fits this mode. More specifically, as one of the largest wildfires on record, engulfing more than 1.2 two million hectares over just two days, it is widely viewed as having been the catalyst for the introduction of a rigorous national fire suppression regime in the US that lasted for much of the twentieth century (van Wagtendonk 2007). While clearly of importance, the focus on the wider implications of such environmental shocks often has led to the neglect of their other indirect effects. In terms of large wildfires, one secondary consequence is the released air pollutants which can travel considerable distance, imposing a substantial burden on human, economic, and environmental systems in areas potentially very far from the actual fire incident (Sapkota 2005; Kollanus 2017). The Great Fire of 1910 was no exception, where subsequent smoke drifted nearly a thousand miles from where the fire had started (Egan 2009, p. 228). Importantly, at the time the potential health effects of wildfire smoke exposure appear to have been partially undervalued and considered only of secondary concern compared to the direct economic consequences, thus likely minimising avoidance behaviour and potentially exacerbating any impact (Burke 2022). In this paper, we set out to explicitly measure the health impact of smoke exposure due to the Great Fire of 1910.

Fire-sourced air pollution and health outcomes have been at the centre of a rich body of literature in modern settings. In this regard, smoke exposure has been linked to a number of short-term effects, including negative physiological (e.g. respiratory and cardiovascular morbidity and mortality) and psychological (e.g. post-traumatic stress disorder, depression) impacts; for extensive reviews see Liu et al. (2015), Chen et al. (2021), Grant and Runkle (2022), Gao (2023). In contrast, there are only a handful of studies that have examined its effect on health in historical contexts in the short term, despite polluted air having been a problem since mediaeval times (Brimblecombe 2011). For example, Beach and Hanlon (2018) use information on local coal use and wind patterns to demonstrate that exposure to air pollution in the 1850 s increased infant mortality rates across England and Wales by between 6 and 8 per cent, respectively. Also, Barreca et al. (2014) show that reductions in the use of bituminous coal for heating between 1945 and 1960 decreased winter all-age mortality, while Clay et al. (2016) provide evidence that being near major coal-fired power plants increased infant mortality.

The negative physical and psychological impacts of pollution can potentially also have important negative long-term consequences, although even in modern settings the evidence is scarce. Exceptions include Deryugina and Reif (2023) and Colmer et al. (2024), who demonstrate that pollution can lead to reduced survival and inequality, respectively. In a historical context, Bailey et al. (2018) find that local coal intensity exposure of WWI enlisted English and Welsh men during childhood had

[1] For example, the 1755 Lisbon earthquake is believed to have led to a change in the cultural and social perception of natural disasters, and thus risk management, across Europe (Mendes-Victor et al. 2009).

longer-term health effects by reducing adulthood height. While this is the only non-modern study that we are aware of, studies of other environmental toxins provide some clues as to possible long-term effects. For instance, Noghanibehambari and Fletcher (2022) studying the effect of in-utero and early-life exposure to soil erosion caused by the Dust Bowl in the 1930 s find that longevity was reduced, especially for children raised in family households, females, and those with lower maternal education, while Frye and Kagy (2023) show that children exposed to waterborne lead in the late nineteenth century had lower later-life incomes, worse occupations, and lower levels of education. Although Heblich et al. (2021) do not look specifically at health, they provide evidence that air pollution during the industrial revolution modified the spatial organisation of cities (i.e. low-skilled workers reside in high-pollution areas) which still persists today, and one would suspect that negative health effects are likely to also have played at least an indirect role in this geographical allocation.

In this paper, we add to the nascent historical literature not only by exploring for the first time the impact of another source of air pollution, namely wildfires, on health, but also by estimating both its short-term effects and long-term implications in terms of socioeconomic outcomes.[2] To this end, we avail of a historical map of the burnt areas of the Great Fire and model the smoke emission and dispersion to reconstruct the wildfire smoke exposure at the county level using meteorological inputs. Arguably constructing air pollution this way allows us to capture the causal effect of fire-sourced smoke exposure, in particular because at the time anticipatory behaviour, as noted above, is likely to have been minimal. We first combine our smoke exposure proxy with digitised weekly under age of five deaths at the county level to assess its short-term, post-natal impacts on excess mortality. To assess more longer-term impacts on economic outcomes on those that survived, we then use matched census data to link boys who in 1910 were under the age of five to their socioeconomic status outcomes 20 and 30 years later and compare the men who were smoke-exposed in their early childhood to men who were not.

Our econometric results suggest an immediate impact of smoke exposure on excess mortality for children under the age of five in the week of the wildfire event. More specifically, the weekly excess mortality rate is 35 per 100,000, which translates into an increase of $\approx$ 119% relative to the average observed mortality rate in 1910 across the entire study area. Furthermore, we find evidence of a negative effect of 8–14% of wildfire-sourced smoke exposure in early childhood on composite later-life socioeconomic status indicators derived from occupational income, education, and additional information on prestige in 1930. In 1940, the results show that men who were smoke-affected in their early childhood encounter a decrease in the occupational income of about 90 US$ per year, which represents a 3.6% reduction compared to the 1940 average occupational income.

[2] One may also want to note that previous studies on other economic consequences of fires in non-modern contexts have focused mainly on urban rather than forest fires. See for example, Hornbeck and Keniston (2017) and Siodla (2015) who study the Great Boston Fire of 1872 and the San Francisco 1906 earthquake, respectively.

The remainder of this paper is structured as follows. Section 2 presents the historical background and the unfolding of the Great Fire of 1910. Section 3 provides details of the data sets used followed by Sect. 4 which explains the empirical framework. In Sect. 5, we present the results and robustness checks. Section 6 concludes.

2 Historical background

The Great Fire of 1910, commonly known as the Big Burn or the Big Blowup, was a catastrophic event that occurred in a rugged and remote geographical region characterised by towering mountains, dense forests, and pristine alpine lakes.[3] Fires, which were oftentimes started by mine operators, coal-burning trains, or lightning strikes, were a common occurrence in the region during the period. While the native population had adapted to the regularity of wildfires, new settlers were largely unfamiliar with their potentially devastating consequences. In the spring and summer of 1910, warm and dry weather conditions were highly anomalous, as shown in Diaz and Swetnam (2013), and the first wildfires were reported as early as April, exacerbating concerns over safety (Kelley et al. 1944; Weigle 1934, p. 166). On 26 July 1910, a lightning storm ignited numerous fires in the Bitterroot Mountains straddling the border between Idaho and Montana and prompted the United States Forest Service (USFS) to request additional firefighters.[4] As the situation worsened in August, all available men were dispatched to the fires. Shortly thereafter, President William Taft authorised the deployment of 2500 military troops, including the 25th Infantry, also known as “Buffalo soldiers” to combat the fires. Along with many others deployed to the towns of Wallace and Avery, Idaho, and to Missoula, Montana, they lacked any experience or training for firefighting (Kelley et al. 1944, p. 168).

On the 20 th of August, strong hurricane-force Palouser winds from the southwest picked up and merged a number of small fires into one large, all-consuming conflagration. The Wallace town Mayor ordered an evacuation by train by midnight that day.[5] The fire was finally extinguished on the 22nd of August when a cold front swept over the Northern Rocky Mountains, bringing steady rain and some early snowfall in the high-altitude alpine reaches. Overall, the fire raging across the states of Idaho, Montana, and Washington shown in Fig. 7 in Appendix A claimed the lives of 87 people, primarily firefighters, burned five towns to the ground, and partly destroyed many others. It burned over 3 million acres ($\approx$ 1.2 million hectares) of forest, an area about 1.5 times larger than Yellowstone National Park (Egan 2009, p. 172).[6]

[3] The information presented in this section is primarily derived from Egan (2009), unless otherwise specified.

[4] The USFS was established in 1905 under the presidency of Theodore Roosevelt, a strong advocate for nature conservation, to manage 150 national forests.

[5] The evacuations were carried out very late and only moved people out of the immediately threatened burn zone. Thus, evacuees were relocated to areas heavily affected by smoke (e.g. Missoula), so evacuees still experienced exposure before, during, and after the evacuation.

[6] There were no children among the recorded casualties.

The subsequent smoke created by the fires was dispersed by winds over vast areas well beyond the burnt regions. For instance, Egan (2009, p. 228, 172) in his book recounting the event notes that "in Denver... a layer of smoke three thousands feet thick settled over the city, the caboose of the Big Burn, nearly a thousand miles from where it started" and that "smoke drifted hundreds of miles from the blowup, into the Dakotas and Colorado and Alberta and Wyoming. It was if a volcanic blast had disgorged the airborne remains of the forested northern Rockies into disparate parts of the United States". The Missoulian (Missoula, Montana) reported on the 21st of August that "the pall of smoke overhanging the town was so dense that the electric lights were turned on at 3 o'clock in the afternoon". One day later, the Great Falls Tribune (Great Falls, Montana) mentions the wildfire smoke a number of times and that "A dense pall of smoke hangs all over western Montana. In Missoula it was as dark as midnight at 5 o'clock, the dense smoke being given a lurid hue, which had all the semblance of the glow of fire, but which was probably due to the sun". For Great Falls, a town that was, similar to Missoula, spared by the flames, the column reads "The atmosphere in this vicinity has been heavily charged with smoke for the past twenty-four hours and at a late hour this morning it was almost impossible to distinguish objects in any portion of the city for a distance of three blocks". Even the cause of a derailed train was partly attributed to poor visibility mentioning that "Train No.2 on the Great Northern was wrecked one and a half miles west of Rudyard... The cause of the wreck was attributed partly to the dense smoke which prevailed last night and today. The sun being entirely obscured and it being only possible to see a few hundred feet ahead".

In considering the effect of smoke from the fires on health, it is important to know how much people may have acted to avoid their exposure to it. In this regard, Egan (2009, p. 141) points out that "people could tolerate the ever-present smoke, though it wasn't good for children and the elderly, made eyes redden and throats scratchy and brought on a ragged cough... They put up with these low-grade tortures because shorter days told them summer was almost over, and they had lived through a humdinger, and soon the rains would come and wash the town clean". More generally, one should note that at the time the nuisance of smoke in general was regarded more in terms of its impact on visibility (Stradling and Thorsheim 1999). For example, city legislation to control urban air pollution was defined accordingly (Goss 1916). Additionally, the scientific evidence of potential negative effects of smoke inhalation during this period was partly (Goss 1916; Benner and O'Connor 1913). Another factor that may have played a role in reducing avoidance behaviour is that people in forest fire-affected regions were so accustomed to periodically seeing smoke that, as Reynolds (1903, p. 28) noted after a study of fires in Marinette County in Wisconsin "that the sight of smoke rolling up provokes little comment".

Finally, one may want to note that the catastrophic Great Fire of 1910 was instrumental in solidifying support for the USFS's fire management mission and led to the establishment of a vigorous fire suppression regime, although the main reasons appear to have been related to the supply of timber as well as the role of forests in flood prevention.[7] In 1935, the 10 o'clock rule was introduced, which mandated that every fire must be extinguished by 10 am the following morning. The mission of relentless fire prevention and suppression was consolidated by the invention of "Smokey Bear" in 1944, a symbol for the joint effort to promote forest fire prevention. According to van Wagtendonk (2007), fire suppression was the only fire policy implemented by the federal land management agencies and it was not until 1974 that the USFS transitioned to a fire management regime allowing some fires ignited by lightning to burn in specific wilderness areas.

3 Data and descriptive statistics

3.1 Burn perimeters

A historical map of the Great Fire of 1910's burn perimeters created by the USFS is accessible from a number of historical sources. We obtain the required data from the newspaper Spokesman Review.[8] The digital raster image is georeferenced in ArcGIS by using the shape file of the US states provided by the US Census bureau.[9] Figure 1a shows how the image of the historical burned area is matched to a spatial reference system using state boundaries. We then delineate each fire scar individually and create a shape file consisting of 176 burned area polygons.

3.2 Smoke modelling

While modern studies on the health impacts of air pollution can use satellite imagery or ground monitoring stations, these options are not available for 1910. Historical research on industrial air pollution has either examined stationary pollution in proximity to its source (Bailey et al. 2018; Beach and Hanlon 2018) or employed an industrial pollution model with contemporary meteorological data (Heblich et al. 2021). Given none of these approaches seem suitable for our study, we use the BlueSky modelling framework which is specifically designed to model the emissions and smoke plume dispersion resulting from different types of fires, such as wildland, agricultural, and prescribed fires. BlueSky operates as a modular framework integrating advanced models and data sets in the fields of fuels, consumption, emissions, meteorology, and air quality within a cohesive structure and allows for multiple options for each stage of the modelling process (Larkin 2009). The sequential modelling steps start with information on the precise fire location

[7] For instance, a leaflet by James Wilson, the United States Secretary of Agriculture, printed in 1900 states the following: "The great annual destruction of forests by fire is an injury to all persons and industries. The welfare of every community is dependent upon a cheap and plentiful supply of timber, and a forest cover is the most effective means of preventing floods and maintaining a regular flow of streams used for irrigation and other useful purposes (Miller and Cohen 2021, p. 8)".

[8] The full newspaper article is available at https://www.spokesman.com/stories/2010/aug/15/1910-fire-region-consumed/.

[9] The shape files and related geographic information are sourced from https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html.

and fuel loads, advancing to fuel consumption, and concluding with smoke emissions and transport.[10]

Figure 2 provides a schematic overview of the Bluesky modelling framework. The darker shaded parts indicate where a value provided by the user is necessary for the models to run or where our input choice deviates from the default settings. The first part, which is shown in the left-hand column in Fig. 2, comprises information on the emissions of the fire. To create the necessary values in the fire information section, we calculate the size in acres as well as the longitude and latitude of the centroid for each of the burned area polygons. In addition, the fire type is specified as "wildfire" as opposed to prescribed or agricultural fire. The fuelbed type is automatically selected based on the location of the fire on a modern map.

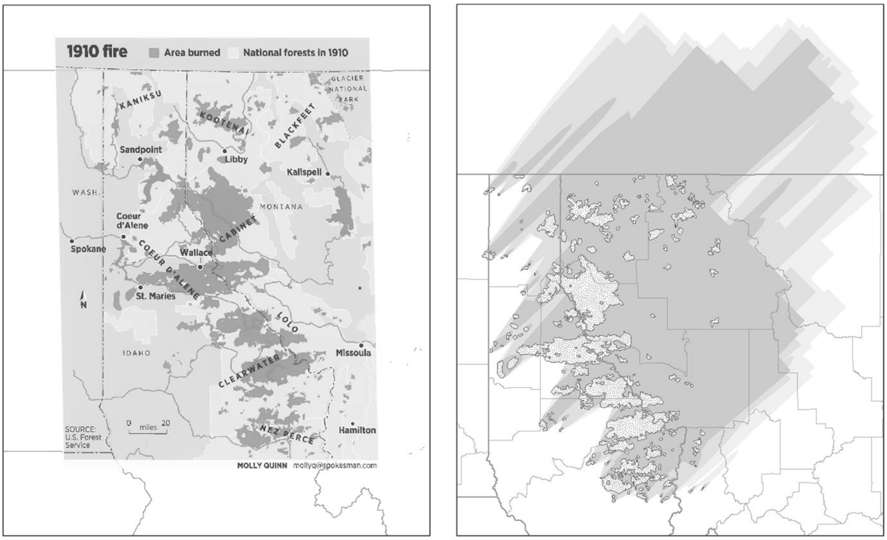
Within the fuel moisture inputs section, the season is specified as "Summer". Regarding the fuel moisture section, the moisture level is set to the most extreme value "Very dry" informed by the work of Diaz and Swetnam (2013) which noted exceptional hot and dry climatic conditions in 1910 that were not matched again until 2012. This choice is reinforced by information from Egan (2009). To maintain clarity and brevity, only the essential model choices are outlined in the main text, although quotes from Egan (2009) are taken as further supporting evidence to specify our input choices for all modelling steps as presented in Table 7 in Appendix A. The default settings regarding the percentage of shrub and canopy consumption remain unchanged. Although the Great Fire lasted for approximately 36 h, the simulation is run for one day as each polygon is modelled individually and likely burned for no more than 24 h. Finally, guided by historical records we selected the flaming combustion phase over the smouldering combustion phase within the timing section.

The output from the smoke emission input modelling generates an emission report that is subsequently used in the second modelling step, the smoke dispersion modelling shown in the middle column of Fig. 2. For the dispersion modelling part, we use the visibility smoke modelling system (VSMOKE), which uses Gaussian plume equations to simulate how smoke particles disperse in the atmosphere under various meteorological conditions. These conditions are carefully chosen to accurately mirror the historical conditions at the time of the Great Fire as documented in historical records.

We select the stability class "Moderately Unstable" and thus deviate from the default setting of "Near Neutral" for two reasons. First, there is evidence of turbulent atmospheric conditions due to the Palouse wind event triggering the wildfire escalation. Second, the Great Fire of 1910 is likely to have created its own weather system that influences atmospheric conditions. Although research on strong convective processes associated with extreme wildfire events, known as pyroconvection, has only evolved recently as noted in Dowdy and Pepler (2018), and therefore the specific identification as such may not have been used in historic reports, there is strong evidence of pyroconvection during the Big Burn.

The wind direction, predominantly from the southwest at 225°, characterises the Palouse winds during the event. While the peak windspeed was reported as "...the

[10] Modelling simulations are conducted using version 3.5.1 of the interactive online platform BlueSky Playground and is accessible under https://tools.airfire.org/playground/v3.5/emissionsinputs.php.



(a) Historical burn perimeters **(b)** Modelled smoke plumes

Fig. 1 Historical map of the burn perimeters and the resulting modelled smoke plumes. *Notes:* (i) the historical map shown in panel (**a**) created by the USFS shows the burn perimeters of the 1910 fires and is georeferenced using ArcGIS; (ii) the county shape file is provided by the Big Ten Academic Alliance Geoportal and shows the historical county boundaries in 1910; (iii) the merged smoke plumes shown in (**b**) are modelled using BlueSky Playground version 3.5.1; (iv) in (**b**) the grey dotted area shows the burn perimeters. The darkest grey-shaded area indicates the area where the hourly peak $PM_{2.5}$ pollution is hazardous ($PM_{2.5}$ > 526 μg/m^3). The medium grey area shows unhealthy hourly peak $PM_{2.5}$ pollution of the values ($PM_{2.5}$ > 130 μg/m^3), and the lightest grey scale denotes moderate hourly peak pollution ($PM_{2.5}$ > 38 μg/m^3)

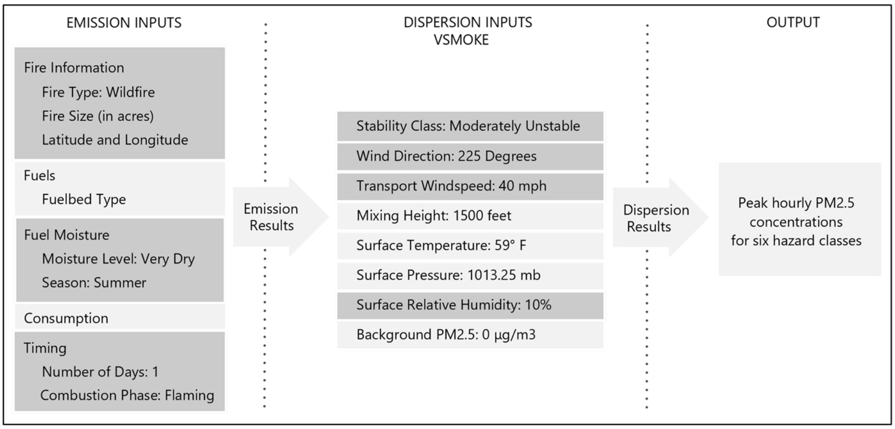


Fig. 2 Schematic overview of the smoke emission and dispersion process using the BlueSky modelling framework. *Notes:* (i) the above model shows the conceptual framework of the BlueSky modelling tool (BlueSky Playground version 3.5) divided into the emission and dispersion inputs; (ii) the darker shaded parts of the framework indicate where value supplied by the user is strictly required for the framework to run or where our input choices deviate from the default settings; (iii) F = Fahrenheit, PM = particulate matter, mb = millibars, μg/m^3 = micro-grams per cubic metre

conscripted air was no longer a Palouser but a firestorm of hurricane-force winds, in excess of eighty miles an hour (Egan 2009, p 156)", it has also been noted that there were windspeeds of 50–60 mph (Egan 2009, p 155, 158). However, since we want to select a value that is arguably more appropriate to capture an average over the 24-hour model simulation, we chose a value of 40 mph. Finally, with a value of 10% for relative humidity, we deviate from the default value of 25% based on climatological records by Diaz and Swetnam (2013), who point out extremely low relative humidity with values around 20% or lower for the affected areas.

The smoke modelling is carried out individually for each of the 176 burned area polygons. The dispersion result indicating the peak hourly $PM_{2.5}$ concentration plume in $\mu g/m^3$ is available in six hazard levels. We merge the smoke plumes from all individual fire polygons considering three levels of pollution: (i) "hazardous", where peak hourly $PM_{2.5}$ concentrations are above 526 $\mu g/m^3$, (ii) "unhealthy", with peak hourly $PM_{2.5}$ concentrations exceeding 130 $\mu g/m^3$, and (iii) "moderate", where the hourly peak $PM_{2.5}$ pollution was at least 38 $\mu g/m^3$. The merged smoke plumes with the three different concentration levels are shown in Fig. 1b. To test the sensitivity to our parameter choices, we run an additional model keeping the wildfire default settings whenever no user input is required (with the exception of wind speed where the default value is 0.5 mph which is unreasonably low for this scenario), i.e. the setting for moisture level is "Dry", the stability class is "Near Neutral", and the surface relative humidity is 25%. These changes do not alter which counties are defined as smoke affected in our estimations as shown in Fig. 8 in Appendix A.

3.3 Population data

We use the anonymised full-count census population data provided by the Integrated Public Use Microdata Series (IPUMS) USA for the decades 1900 and 1910 (Ruggles et al. 2024). Note that the 1900 census data were collected on the 1st of June 1900 (ISO week 22) and the 1910 census data on the 15th April 1910 (ISO week 15).[11] Accounting for boundary changes in this period, which are explained in Sect. A.1 in Appendix A, the weekly population numbers are linearly interpolated using the two censuses and a weekly county-level population panel data set from 1905 to 1910 is created.[12]

3.4 Mortality data

We obtain the county-level mortality data for the years 1905 to 1910 from the private genealogy company Ancestry.com LLC using their online products of comprehensive digitised death records including i.a., Ancestry.com, and Find a Grave. We run separate searches for deaths for each of the years per county for the three

[11] The information on the exact census dates is given at https://www.census.gov/history/www/through_the_decades/overview/.

[12] The adjustment may introduce measurement error. However, since the boundary changes are arguably unrelated to smoke exposure, this could only lead to a bias towards zero of our estimates and would not invalidate our results.

fire-affected states of Idaho, Montana, and the eastern part of Washington. The search results on Ancestry.com draw on a number of digitised data sources, such as state and county records, newspapers' obituary sections, church records, and gravestones. Table 9 in Appendix A provides an overview of the data sources on Ancestry.com used for our specific search.

The detailed steps of the cleaning procedure are outlined in Table 10 in Appendix A. Our final data comprise the deaths from 1905 to 1910 for the 70 sample counties and include 7,685 individual mortality records for children under the age of five.[13] We aggregate the individual death records to weekly death counts using the ISO week date system as each week consists of 7 days. Moreover, we adjust the weekly death counts to county boundary changes during this period in a similar manner to that previously described for the population data.

3.5 Construction of the excess mortality rate

Estimating excess mortality has emerged as a progressively effective method for quantifying the impact of an event (Acosta and Irizarry 2022). The concept of excess mortality has been applied to a wide range of different events, including more recently to the COVID-19 pandemic (Karlinsky and Kobak 2021; Msemburi 2023). In the realm of natural disasters, Santos-Burgoa (2018) estimate the excess deaths related to Hurricane Maria in Puerto Rico, while Morita (2017) studies the indirect excess mortality risk associated with the 2011 triple disaster in Fukushima, Japan. For wildfires specifically, excess mortality risk has been studied by Hänninen (2009) in Finland related to East European wildfires and by Kochi et al. (2012) for the 2003 southern Californian wildfires.

Excess mortality can be defined as the additional deaths that occur in a given period of time due to a health event relative to the deaths that would normally have occurred in its absence. Since the counterfactual, i.e. the number of deaths in the absence of the health event cannot be observed, a common approach to estimate excess mortality is to subtract the baseline mortality from the observed mortality. This baseline is often based on observations for the same region prior to the event. As implemented by many health monitoring institutions worldwide (e.g. European mortality monitoring activity (EuroMOMO) and the United Kingdom Office for Health Improvement and Disparities), we use a reference period of five years prior to the event, i.e. we calculate the weekly moving mortality average for children under five years of age from 1905 to 1909 for each county to give us our baseline.

Our context is the impact of the Great Fire of 1910 measuring excess mortality at the regional level. Since the study area is rural and sparsely populated, and the temporal resolution is relatively low, i.e. weekly, it is important to take the population size into account. Therefore, we use death rates, calculated as the number of deaths per 100,000 of the population. In addition, since mortality rates are affected by seasonality throughout the year, we calculate the week-specific mortality rate for each ISO week of the year. Weekly excess mortality rates are derived from a three step

[13] As described in Table 10 in Appendix A, we exclude children aged zero from the sample as they cannot confidently be detected as duplicates. For simplicity, we refer to children under five in the text.

procedure. First, we calculate the baseline mortality rate for our reference period. As death counts are volatile, our baseline mortality rate is smoothed using each specific ISO week ± 1 week as shown in Eq. (1)[14]:

$$BMR_w = \frac{1}{15} \sum_{t=1905}^{1909} \sum_{w=-1}^{1} MR_{wt}, \tag{1}$$

where *BMR* represents the baseline mortality rate for children under the age of five in ISO week w, t denotes the year, and MR_{wt} is the mortality rate in week w in year t. Additionally, we conduct robustness checks through a multiverse analysis, as detailed in Sect. 5.3.

The baseline mortality rate estimations for children under the age of five range from approximately 9 to 16 weekly deaths per 100,000 as an average for all the counties in our sample. Figure 9 in Appendix A shows the seasonal trend over a 52-week year. The highest mortality rates are observed in the summer and mortality tends to decrease in spring and autumn (shaded in grey). The mean value is 12 which is shown in the first row of Table 1. We benchmark our estimates to the official 1910 mortality statistics reports by the Department of Commerce and Labour at the Bureau of the Census. The reported all-age annual mortality rates for 1905–1909 for rural regions in the registration area are equivalent to approximately 25 to 27 weekly deaths per 100,000 (Department of Commerce and Labor 1912). Thus, our mortality rate estimate for the age group under five (additionally taking account of seasonality) is approximately 54% lower than the official records for all-age mortality. This provides validation for the completeness of the comprehensive mortality data sources used in this study since we naturally expect a lower value for the sub-population under investigation. Yet, although the age group under five makes up around 8–9% of the population (see Sect. 4.2) it seems sensible that due to high child mortality at the time this vulnerable population group has much higher than proportional mortality rates.

Second, we calculate the county-level observed mortality rate of each ISO week in 1910 shown in the second row of Table 1. The observed average weekly mortality rate for children under the age of five in 1910 is 16 per 100,000 and on average, higher than the baseline rate of 12 that is derived from the reference period. This might be because we are capturing more digitised death records in 1910 than in the previous years potentially driven by the fact that the state of Washington became part of the mortality statistics registration area in 1908 and Montana in 1910.

Third, we calculate the county-level weekly excess mortality rates for all ISO weeks of 1910 by subtracting the baseline mortality rates derived from the reference period 1905–1909 from the observed weekly mortality rates in 1910. The third row of Table 1 shows the excess mortality rate descriptive statistics for all our observations, i.e. 52 weeks for 70 counties.

[14] Using a smoothing approach is standard in widely applied excess mortality algorithms used by public health offices such as the Farrington and Noufaily algorithms described in Noufaily (2012).

Table 1 Descriptive statistics of weekly baseline, observed, and excess mortality rates for children per 100,000 under the age of five in 1910

	Min	Mean	SD	Median	Max	*N*
BMR Age < 5	0	12	14	8	152	3640
OMR Age < 5	0	16	51	0	855	3640
EMR Age < 5	− 152	4	51	− 4	855	3640

Notes: (i) *SD* standard deviation; (ii) BMR indicates the baseline mortality rate from 1905 to 1909, OMR denotes the observed mortality rate in 1910, and EMR is the excess mortality rate in 1910; (iii) the population data are compiled from the 1910 US full-count census provided by the Integrated Public Use Microdata Series (IPUMS) USA; (iv) the mortality data are retrieved from the genealogy company Ancestry.com; (v) the maximum OMR value of 855 stems from Granite County, Montana, in week 22 with two recorded deaths for children under five for a population of 233.8 for children under the age of five

3.6 Census crosswalks

For the analysis of the long-term effect, we use the IPUMS full-count household and individual data for the years 1910, 1930, and 1940 (Ruggles et al. 2024). In order to link individuals over time, we use the crosswalk files provided by the Census Linking Project implementing the Abramitzky, Boustan, and Eriksson (ABE) exact conservative algorithm (Abramitzky et al. 2022a, b). The links are undertaken based on variables that are expected to remain constant over time, typically the birth year, name, gender, and county or state of birth; see Abramitzky et al. (2021) for the stepwise procedure. As demonstrated in Abramitzky et al. (2021), automated matching approaches significantly reduce the rate of false positives. In pursuit of the highest precision and to minimise false positive matches further, we select the most stringent algorithm available, which requires individuals to be unique by name within two years to qualify as a match. Note that in the crosswalk files only males can be linked because females are harder to track given many change their last names after marriage. Furthermore, IPUMS does not provide the date of birth, and thus, we are unable to study in-utero exposure and therefore focus on post-natal effects.

For 1910, we use all individuals of the full-count census to create a number of control variables at the county level, including average age, percentage of farm households, and the economic structure captured as the percentage of workers in different industries. The full list and descriptive statistics of the county-level variables are shown in Table 11 in Appendix A. Furthermore, we extract the boys who were under the age of five in 1910, including their corresponding parental information. We link the boys with their household characteristics in 1910 using the household serial number and merge our variables of interest with the crosswalk files for 1930 and 1940.

At the individual level, we construct a race indicator to identify non-white individuals.[15] Non-nativity is defined as one if at least one parent is born outside the US. At the household level, we include family size in terms of the number of members,

[15] Note that the population share of natives in the study area was about 1%.

the number of families within a household, and indicator variables for whether it is a family-, farm-, urban-, or mortgage-paying household. The parental characteristics on the mothers' side are limited as women are usually not part of the labour force although we do include the age and nativity of the mother. For the child's father, we obtain numeric information regarding their age, occupational education, and earnings scores. Indicator variables on being employed as well as the industries where they work are created, where industry codes are classified into 12 categories based on the 1950 Census Bureau industrial classification system. Table 12 in Appendix A shows the descriptive statistics for the 7801 boys that we were able to link from 1910 to 1930. Linking approximately 7800 (9000) boys in 1930 (1940) out of about 98,400 children under the age of five in 1910 is equivalent to a matching rate of 15.9% (18.3%) assuming 50% boys.[16] Note that the linked sample for 1940 is slightly larger likely due to more comprehensive record taking.

3.7 Socioeconomic status

The outcome variables of interest in the second part are the socioeconomic status measures in 1930 and 1940.[17] Monetary wage and salary data are not available until 1940, and we therefore use the two available composite measures: the Duncan Socioeconomic Index (SEI) and the Nam–Powers–Boyd Occupational Status Score (NPBOSS), which are both derived from the 1950 occupational classification scheme. The SEI is a metric on a scale from 0 to 96 constructed by regressing prestige ratings from the 1947 National Opinion Research Centre survey on occupational education and occupational income and is referred to as a "socioeconomically predicted prestige" scale (Duncan 1961). The NPBOSS is based upon median earnings and median educational attainment associated with each category in the 1950 occupational classification scheme. It is on a scale from 0 to 100 and can be interpreted as the percentage of persons who are in occupations having lower combined levels of education and earnings than the respondent and can be referred to as a "pure socioeconomic" scale (Nam and Boyd 2004).

While the composite measures are useful as a proxy to capture multiple dimensions defining occupational standings, they have faced criticism particularly regarding the study of social mobility due to lack of consistent specification (Hauser and Warren 1997) and a misguided rating of women's occupations (England 1979). However, since in this study we only link men and do not compare a single individual's occupational standings over time, the aforementioned issues are arguably inconsequential to our analysis. Furthermore, we are comparing men in a treatment and comparison group at the same point in time, and thus, variation in definition over time does not affect our estimates. Nevertheless, we additionally include

[16] Our match rates, which are slightly below the approximately 20% reported in previous studies that link samples between censuses (Abramitzky et al. 2012; Collins and Wanamaker 2014; Long and Ferrie 2013), result from our decision to employ the most stringent matching criteria using conservative matching. Implementing the less strict ABE standard matching approach, our matching rate aligns with the established 20% benchmark.

[17] Note that in this study socioeconomic status and occupational standings are used interchangeably.

Table 2 Descriptive statistics of the socioeconomic status indices of the linked men in 1930 and 1940

	Min	Mean	SD	Median	Max	*N*
1930						
Duncan Socioeconomic Index (SEI)	3.0	22.2	19.5	15.0	96.0	7801
Nam–Powers–Boyd Occ. Status Score (NPBOSS)	3.6	35.2	26.4	25.1	100.0	7738
Occ. Income Score [in 100 s US $] (OCC.INC)	3.0	19.3	9.5	20.0	80.0	7801
Occ. Earnings Score (OCC.ER)	0.6	35.5	28.1	39.7	100.0	7738
Occ. Education Score (OCC.ED)	0.3	9.7	15.5	3.3	96.0	7712
1940						
Duncan Socioeconomic Index (SEI)	3.0	31.6	24.2	18.0	96.0	9042
Nam–Powers–Boyd Occ. Status Score (NPBOSS)	3.6	48.5	28.2	48.7	100.0	9162
Occ. Income Score [in 100 s US $] (OCC.INC)	3.0	24.9	11.1	24.0	80.0	9184
Occ. Earnings Score (OCC.ER)	0.6	49.2	29.8	52.6	100.0	9162
Occ. Education Score (OCC.ED)	1.0	15.8	21.9	4.6	93.8	9139

Notes: (i) *Occ.* occupational, *SD* standard deviation; (ii) the variables are obtained from the Integrated Public Use Microdata Series USA 1930 and 1940 full-count individual censuses; (iii) the 1940 links are newly matched men and the number of links is likely higher than in 1930 due to better record taking

dimension-specific measures of occupational income (OCC.INC), earnings (OCC.ER), and education (OCC.ED) also derived from the 1950 occupational classification scheme. The descriptive statistics of the socioeconomic status indices in 1930 and 1940 are shown in Table 2.

4 Empirical framework

4.1 Identification strategy

Natural hazards like earthquakes or hurricanes typically affect specific high-risk regions, but the exact locations impacted within these areas can be somewhat random. In contrast, wildfires tend to occur in non-random patterns and local occurrence may be influenced by various economic factors rather than being entirely exogenous events. Reasons for this include that wildfire incidence is dependent on anthropogenic factors such as land-use changes (e.g. deforestation) or land and fire management policies that are potentially correlated with health and socioeconomic characteristics. Further endogeneity concerns arise from the fact that many fires start directly due to human activity either through negligence or intentional actions. In the context of the Great Fire of 1910, some of the ignitions occurred due to sparks flying off coal-burning trains and it is speculated that some of the fires were set deliberately for political and economic reasons.[18] Thus, wildfire occurrence may be correlated with a number of unobserved economic dimensions or behavioural

[18] Source from Professor J.E. Kirkwood's notes on a summer spent in the woods in 1910, Records of the USFS, Region One headquarters, Missoula.

patterns that potentially also affect child mortality or socioeconomic outcomes, and not accounting for such unobserved factors would lead to biased estimates.

To address these endogeneity concerns, we model the smoke plume utilising meteorological inputs, such as wind direction and transport wind speed, which arguably induces exogenous variation in smoke exposure that can be leveraged to estimate the causal effect of smoke pollution on the counties' excess mortality rates and later-life socioeconomic status. The identification strategy of using wind direction to estimate the impacts of fire-sourced air pollution on health has been applied in various contexts. For example, Rangel and Vogl (2019) exploit daily changes in agricultural fire location and wind direction to relate in-utero smoke exposure to health at birth outcomes in the sugar-growing region of the Brazilian state of São Paulo. Furthermore, Rocha and Sant'Anna (2022) employ an instrumental variable strategy combining the monthly variation in wind direction in surrounding municipalities to estimate the effect of deforestation-related smoke pollution on morbidity and mortality for municipalities in the Brazilian Amazon. Finally, Pullabhotla and Souza (2022) use daily data on wind direction to study the effect of agricultural fires on hypertension risk in India.

As our modelling of the smoke due to the wildfires is based on a number of meteorological factors, it can be considered strictly exogenous since these are unlikely to have been anticipated (e.g. no endogenous selection into treatment). To also exclude the possibility that the treatment and the comparison group were nevertheless on different pathways regarding their excess mortality rates before the event, we also test pre-treatment differences of the two groups by including leads of the treatment variable. Furthermore, we are controlling for potential direct economic effects of the fire in a similar manner as for smoke exposure using an indicator variable that is one if a county comprises burned area and zero otherwise. Our approach arguably leaves us with capturing the exogenous residual health effects of air pollution.[19]

For the short-term impacts, we analyse the effect of wildfire smoke pollution on excess mortality implementing a Difference-in-Differences design. The central idea is that the excess mortality rates of a population that is smoke-affected (treatment group) would have evolved in a similar manner in the absence of the smoke as the population that was unaffected by the event (comparison group). Assuming that this key assumption holds, identification of the effect of smoke pollution on health relies upon using the comparison group for the unobservable counterfactual outcome in the absence of the event. One might worry about contamination from potential changes in mortality rates due to other period factors (e.g. epidemics). However, those are unlikely to be related to smoke exposure and are thus not of concern. Moreover, these potentially confounding shocks or common trends would be picked up by the week fixed effects assuming they affect the entire study area.

In the estimation of long-term effects of smoke exposure on later-life occupational standings, we link boys who are under the age of five in 1910 over time and assess their socioeconomic outcomes 20 and 30 years later, i.e. in 1930 and 1940, respectively. Since our data are inherently cross sectional, not allowing us to control

[19] We additionally run robustness checks excluding the fire-affected counties completely from the treatment group.

for all possible confounding county and individual level factors, we control for a large number of individual, parental, household, and county-level characteristics that may affect later-life socioeconomic status in order to isolate the remaining variation attributable to smoke exposure. Thus, for the socioeconomic regressions we assume causal identification strategy conditional on these controls. In order to address concerns over migration, we estimate a probit model where an indicator variable denoting whether an individual moved counties between censuses is regressed on smoke exposure, as well as on all the other variables used in the main estimation for the long-term effects. We find that the likelihood of an individual to move counties is unrelated to smoke exposure in 1910 in all the following census years 1920, 1930, and 1940.[20]

4.2 Treatment and comparison group

The categorisation of counties into a treatment group consisting of 14 counties and the comparison group of 56 counties is shown in Fig. 3, where we additionally include the burned area to jointly visualise the source of the fires.[21] Ideally we would include a measure of smoke intensity additionally to the binary treatment classification. Yet, the output from our smoke model identifies the maximum hourly pollution levels in terms of hazard classes. Due to the scale of this event, a significant number of counties fall into the highest hazard category, which unfortunately limits our ability to distinguish between varying degrees of smoke intensity.

In terms of population, in 1910 about 191,000 people were living in the treatment area and 937,000 in the comparison area. Approximately 15,300 (8%) and 83,100 (8.9%) are children under the age of five in the treatment and comparison group, respectively.

Table 3 shows the balance tests across treatment and comparison counties for a number of variables in terms of their eight month average before the Great Fire of 1910. Accordingly, the average population per county is 13,533 and 16,498 in the treatment and comparison areas, respectively. The number of children under the age of five is slightly higher in the comparison group per county. As for the calculated mortality rates, the baseline mortality rates and the observed mortality rates are on average slightly larger in the treatment group, and the excess mortality rates are slightly smaller in the treatment group prior to the event. However, reassuringly none of the variables of interest are significantly different pre-treatment as shown by the t-statistic and corresponding *p* values.[22]

[20] See Table 13 in Appendix A.

[21] Excluded from the treatment group are the two counties that had less than 1% smoke coverage.

[22] Given the long-term estimation is inherently cross-sectional we do not consider a balance-type table necessary as the comparison variables are effectively accounted for within our analysis framework. Yet, we provide a balance-type table for the long-run estimations (see Table 14 in Appendix A) as this was of concern to some readers.

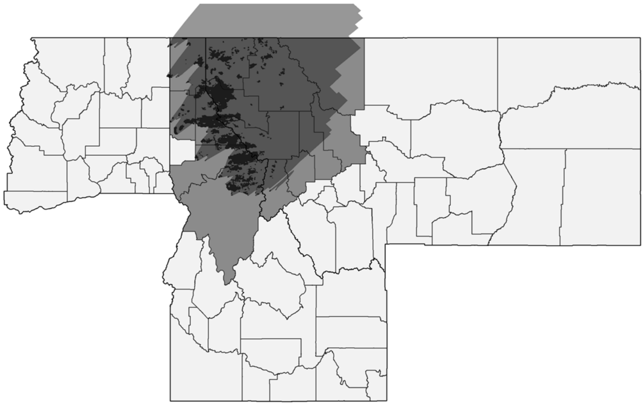


Fig. 3 Treatment and comparison group for smoke-affected counties. *Notes:* (i) a county is classified as smoke affected if any part of the county was exposed to moderate hourly peak pollution ($PM_{2.5}$ > 38 μg/m^3); (ii) the county shape file shows the historical county boundaries provided by The Big Ten Academic Alliance Geoportal; (iii) the darkest grey-shaded area shows the modelled moderate hazard smoke plume employing the BlueSky smoke modelling framework and the black shaded area indicates the burn perimeters of the fire

Table 3 Balance table showing the county-level average population and mortality rates of children under the age of five for the 32 ISO calendar weeks (≈ 8 months) before the fire

	Treatment	Comparison	Difference	t-stat	*p* value
Total population	13,533	16,498	2966	0.55	0.59
Population < 5	1081	1463	382	0.95	0.35
BMR < 5 per 100,000	12.7	11.7	− 0.9	− 0.37	0.71
OMR < 5 per 100,000	17.1	16.7	− 0.4	− 0.11	0.91
EMR < 5 per 100,000	4.4	5.0	0.5	0.16	0.88

Notes: (i) the population variables are obtained from the Integrated Public Use Microdata Series USA 1900 and 1910 full-count individual censuses; (ii) the mortality data are retrieved from the genealogy company Ancestry.com; (iii) BMR denotes the weekly baseline mortality rate per 100,000 and is derived by taking week-specific smoothed averages from 1905–1909; OMR stands for the average observed weekly mortality rate per 100,000 in the ISO calendar weeks 1–32; EMR indicates the weekly excess mortality rate per 100,000 in the ISO calendar weeks 1–32 and is calculated by subtracting the weekly baseline mortality rate from the weekly observed mortality rate

4.3 Econometric specification

We estimate a dynamic two-way fixed effects Difference-in-Differences model for the short-term effects shown in Eq. (2) including 8 pre- and post-event periods (≈ 2 months), i.e. leads and lags of the treatment variable of the wildfire event which is denoted as week 0:

$$EMR_{it} = \sum_{\substack{k=-8 \\ k \neq -1}}^{8} \beta_k \times \mathbb{1}\{t=k\} \times S_i + \sum_{\substack{k=-8 \\ k \neq -1}}^{8} \gamma_k \times \mathbb{1}\{t=k\} \times BA_i + \mu_i + \lambda_t + \varepsilon_{it}, \tag{2}$$

where EMR_{it} is the excess mortality rate in county i and week t, and the week before the event is omitted from the equation to normalise the estimates of β_{-1} and γ_{-1} to 0 in order to estimate the effects relative to the reference period. S_i represents the treatment group indicator and is equal to 1 if county i is smoke-affected in week t, and 0 otherwise. BA_i indicates any burned area as 1 or 0 otherwise. County fixed effects, μ_i, account for county-specific time-invariant characteristics and week fixed effects, λ_t, capture common shocks that might potentially affect our study region at large. ε_{it} is the error component. The coefficients of interest are β_k for $k \geq 0$ (lags) which capture the effect of smoke-exposed in post-event period k relative to the pre-event week $-$ 1. The coefficients of the leads, i.e. β_k ($k < 0$), can be interpreted as pre-event differences in excess mortality rates between treatment and comparison groups.

Error terms ε_{it} are clustered at the county level due to the possibility of persistent correlations between idiosyncratic disturbances within counties on a weekly basis. Thus, we allow for serial correlation within the cross-sectional units over time. Given that the treatment is captured at the county level this is consistent with recent work by Abadie et al. (2023) on how to appropriately cluster the error term.

For the estimation of the long-term effect of smoke exposure on later-life socioeconomic outcomes, we estimate an Ordinary Least Squares (OLS) cross-sectional regression as specified in Eq. (3):

$$\begin{aligned} SES^d_{ic} = {} & \beta_1 S_{ic} + \beta_2 BA_{ic} + IND_i \boldsymbol{\beta}_3 + PAR_{ic} \boldsymbol{\beta}_4 + HH_{ic} \boldsymbol{\beta}_5 \\ & + CTY_{ic} \boldsymbol{\beta}_6 + STATE_{ic} \boldsymbol{\beta}_7 + \varepsilon_i, \qquad d = [1930, 1940], \end{aligned} \tag{3}$$

where SES^d_{ic} indicates the socioeconomic status outcome variable in decade d of individual i who resided in county c in 1910. We study the decades 1930 and 1940. S_{ic} is an indicator variable that is equal to 1 if individual i's county of residence c in 1910 was smoke affected and 0 otherwise. In a similar manner, BA_{ic} represents an indicator variable that is equal to 1 if individual i's county of residence c in 1910 comprised some burned area and 0 otherwise. Furthermore, IND_i, PAR_{ic}, and HH_{ic} are vectors of individual, parental, and household characteristics of individual i who resided in county c in 1910, respectively. The variables contained in these vectors are shown in Table 12 in Appendix A. CTY_{ic} is a vector denoting the characteristics of county c in which individual i resided in 1910. The complete list of the variables included at the county level is shown in Table 11 in Appendix A. Finally, $STATE_{ic}$ is a vector of indicator variables for each state that individual i's county of residence c in 1910 belongs to and ε_i indicates the error term. Standard errors are again clustered at the county level following the similar reasoning as described for the short-term analysis.

5 Results and discussion

5.1 Short-term excess mortality

The point estimates and confidence intervals of Eq. (2) for the 8 pre- and post-event weeks of the Great Fire are shown graphically in Fig. 4. Reassuringly none of the leads are statistically significant, indicating that there was no difference in the excess mortality rate between the treatment and comparison group prior to the event, and hence, as expected, there were no anticipation effects.[23] The results show a positive effect of smoke exposure on the excess mortality rate for children under the age of five in the week of the wildfire, i.e. week 0, but no such effect in the 8 weeks following the event.

Table 4 presents the corresponding regression table resulting from Eq. (2). In the interest of brevity, a condensed version of the estimated lagged coefficients is presented, where the full table showing all lags for week 0 to week 8 is provided in Table 15 in Appendix A. In Column (1), we show the estimates of only including county-fixed effects, while Column (2) presents results of additionally controlling for week fixed effects. Accordingly, only accounting for time-invariant county unobservables implies that smoke exposure had no impact on mortality of under five year olds. In contrast, also allowing for common time specific factors indicates an excess mortality rate in the treatment counties in the week of the fire of 64.4 per 100,000. However, additionally including fire exposure in Column (3) to also capture the direct impact of the wildfires, which are the estimates corresponding to Fig. 4, reduces the point estimate by about 46% per cent. The estimated coefficient suggests that smoke exposure due to the wildfire increased the excess mortality rate by 35 per 100,000 for children under the age of five in the week of the event.[24]

Comparing our estimated excess mortality rate of 35 per 100,000 to the observed weekly mortality rate of 16 per 100,000 over the entire year of 1910, as taken from Table 1, suggests a 119% increase in excess mortality in the week of the fire due to smoke exposure. This immediate impact on excess mortality is in line with findings provided by Johnston et al. (2011) who study all-cause non-accidental mortality due to bush fires and dust storms from 1997 to 2004 in Sydney and report a same-day increase in mortality controlling for temperature for all age groups. Moreover, Doubleday (2020) assess non-traumatic mortality associated with wildfire smoke exposure from 2006 to 2017 in Washington State and report that previous-day smoke exposure poses the highest mortality risk and that it diminishes rapidly within two days.

[23] Similarly, there are no pre– treatment differences for the burned area indicator variable as shown in Fig. 10.

[24] As demonstrated in Eq. (2), we account for the area affected by fires to mitigate potential direct effects on our findings. Nevertheless, to further address concerns over potential direct effects, we re-estimate Eq. (2) while excluding fire-affected counties. The result yields a point estimate of 20 excess deaths per 100,000 children, as detailed in Table 16 in Appendix A. However, this approach is not adopted as our primary estimation strategy due to two key limitations: first, it precludes a significant portion of smoke-affected counties from our analysis, and second, it leaves us with a markedly reduced number of treatment units.

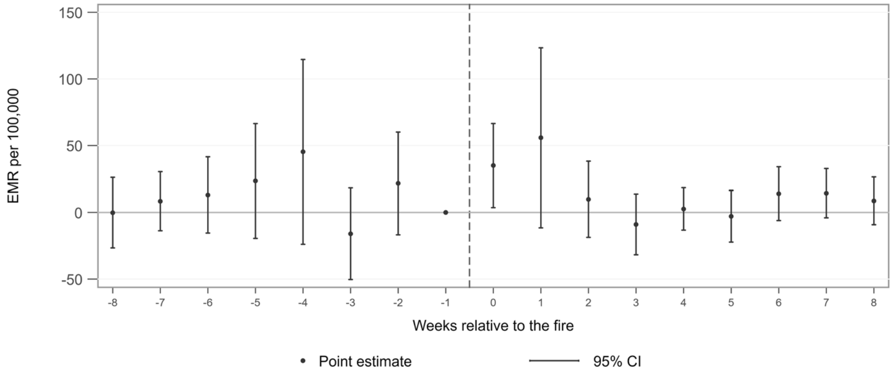


Fig. 4 Difference-in-Differences point estimates and 95% confidence intervals of smoke exposure on excess mortality of children under the age of five. *Notes:* (i) EMR denotes the excess mortality rate and CI indicates the 95% confidence interval; (ii) the mortality data are obtained from the genealogy company Ancestry.com and the population data are retrieved from the 1910 US census

Table 4 Difference-in-Differences regression results of smoke exposure on the excess mortality rate of children under the age of five (weeks 0–4)

	Excess mortality rate		
	(1)	(2)	(3)
Smoke event	51.1	64.4*	35.0*
	(27.8)	(29.3)	(15.8)
1 week after event	48.3	45.7	55.9
	(30.1)	(33.7)	(33.8)
2 weeks after event	22.4	28.8	9.8
	(23.4)	(26.6)	(14.3)
3 weeks after event	2.0	9.7	− 9.0
	(9.6)	(13.3)	(11.4)
4 weeks after event	− 2.1	8.3	2.6
	(9.8)	(12.8)	(7.9)
County fixed effects	✓	✓	✓
Week fixed effects		✓	✓
Burned area			✓
R^2	0.09	0.11	0.11
N	1190	1190	1190

Notes: (i) stars indicate significance according to $*p < 0.05$, $**p < 0.01$, $***p < 0.001$; (ii) the table shows the coefficients of the Difference-in-Differences estimation as stated in Eq. (2); (iii) the sample includes 70 counties; (iv) Burned area indicates whether a county comprised burned area; (v) standard errors are clustered at the county level; (vi) the population data are compiled from the 1910 US full-count census provided by the Integrated Public Use Microdata Series (IPUMS) USA and the mortality data are retrieved from the genealogy company Ancestry.com

Table 5 Regression of individual socioeconomic status outcomes in 1930 on smoke exposure in early childhood

	Composite		Single dimension		
	(1)	(2)	(3)	(4)	(5)
	SEI	NPBOSS	OCC.INC	OCC.ER	OCC.ED
Smoke	− 3.0***	− 2.7*	− 0.7	− 1.8	− 0.8
	(0.8)	(1.2)	(0.5)	(1.4)	(0.6)
Burned area	3.1*	1.0	− 0.1	− 1.6	0.6
	(1.2)	(1.6)	(0.6)	(1.8)	(0.9)
Individual					
Non-white	− 8.7***	− 17.2***	− 6.1***	− 18.2***	− 4.3***
	(1.7)	(2.6)	(0.9)	(3.2)	(1.1)
Non-American born parent	− 0.6	− 0.9	− 0.2	− 0.8	− 0.3
	(0.7)	(1.0)	(0.4)	(1.1)	(0.6)
Household (1910)					
Family size	− 0.8***	− 0.9***	− 0.2***	− 0.6***	− 0.5***
	(0.1)	(0.2)	(0.1)	(0.2)	(0.1)
Families in household	0.3	0.2	− 0.0	− 0.1	0.1
	(0.3)	(0.4)	(0.2)	(0.5)	(0.3)
Non-family household	2.1	3.1	1.0	4.7	2.6
	(3.7)	(5.6)	(2.2)	(6.7)	(3.7)
Urban household	4.9***	5.0***	1.3**	3.2*	2.5***
	(0.7)	(1.0)	(0.4)	(1.3)	(0.6)
Farm household	− 2.3**	− 4.6***	− 1.6***	− 4.8***	− 1.0
	(0.8)	(1.2)	(0.4)	(1.2)	(0.6)
Paying mortgage	0.8*	0.9	0.2	0.7	0.7*
	(0.4)	(0.5)	(0.2)	(0.6)	(0.3)
Parents (1910)					
Mother: Age	0.3***	0.4***	0.1***	0.3***	0.2***
	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)
Mother: Non-American born parent	− 0.3	− 0.6	− 0.4	− 1.1	− 0.2
	(0.4)	(0.6)	(0.2)	(0.7)	(0.4)
Father: Age	− 0.1	− 0.1	− 0.0	− 0.1	− 0.0
	(0.0)	(0.0)	(0.0)	(0.1)	(0.0)
Father: Non-American born parent	0.5	− 0.5	− 0.4	− 1.3	0.1
	(0.5)	(0.7)	(0.3)	(0.8)	(0.4)
Father: Education score	0.1**	0.1	0.0	− 0.0	0.1**
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Earnings score	0.1***	0.1***	0.0***	0.1***	0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Unemployed	− 1.4	− 1.5	− 0.4	− 1.2	− 1.6**
	(0.8)	(1.1)	(0.4)	(1.1)	(0.6)

Table 5 (continued)

	Composite		Single dimension		
	(1)	(2)	(3)	(4)	(5)
	SEI	NPBOSS	OCC.INC	OCC.ER	OCC.ED
Controls					
Industry father	✓	✓	✓	✓	✓
County characteristics	✓	✓	✓	✓	✓
State indicator	✓	✓	✓	✓	✓
R^2	0.11	0.14	0.13	0.13	0.06
N	7801	7738	7801	7738	7712

Notes: (i) stars indicate significance according to $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; (ii) this table shows the results of the Ordinary Least Squares regression shown in Eq. (3) estimating the effect of wildfire smoke exposure in early childhood on later-life socioeconomic status conditional on controls; (iii) the data are obtained from the Integrated Public Use Microdata Series full-count censuses 1910 and 1930; (iv) the composite socioeconomic status measure SEI stands for the Duncan Socioeconomic Index, and NPBOSS is the Nam–Powers–Boyd Occupational Status Score; (v) the single dimension socioeconomic status measures are occupational income (OCC.INC), earnings (OCC.ER), and education (OCC.ED)

5.2 Long-term socioeconomic status

The regression results for the 1930 census linked estimations of Eq. (3) are shown in Table 5. The findings indicate a negative coefficient on both the composite and single dimension indicators. Yet, only the composite measures show a significant negative impact of smoke exposure on the Duncan Socioeconomic Index (0.1% level) and on the Nam–Powers–Boyd Occupational Status Score (5% level). This suggests that there is no single dimension that is the main driver of the negative result of the composite measures of socioeconomic status. The point estimates suggest that the Duncan Socioeconomic Index is 3 points lower for men who were smoke-exposed in early childhood compared to non-exposed men (Column (1)). This translates into a 13.5% decrease relative to the 1930 sample mean of 22.2 (Table 2). The estimation for the Nam–Powers–Boyd Occupational Status Score shown in Column (2) indicates that smoke-exposed men also rank on average 2.7 points lower than non-exposed men, which translates to a decrease of 7.7% relative to the mean of 35.2 points shown in Table 2.[25]

In terms of the other controls, at the individual level non-whites have lower socioeconomic status outcomes than whites. Moreover, growing up in a larger family or on a farm is associated with lower later-life occupational standings, while growing up in an urban household is linked to better performance in later-life socioeconomic

[25] One may be concerned that the boys that survived the wildfire-sourced air pollution are systematically different (e.g. healthier, stronger, richer) than the non-smoke-affected boys and thus, that our sample is characterised by a "survival bias". Assuming this was the case, our results are arguably an underestimate of the true negative effect on the later-life socioeconomic status of the boys who were smoke-exposed.

Table 6 Regression of individual socioeconomic outcomes in 1940 on smoke exposure in early childhood

	Composite		Single dimension		
	(1)	(2)	(3)	(4)	(5)
	SEI	NPBOSS	OCC.INC	OCC.ER	OCC.ED
Smoke	0.1	− 1.1	− 0.9*	− 2.2	1.0
	(1.0)	(1.3)	(0.4)	(1.3)	(1.0)
Burned area	0.6	1.2	0.9	1.8	0.3
	(1.3)	(2.0)	(0.6)	(2.1)	(1.2)
Individual					
Non-white	− 11.3***	− 14.4***	− 3.7**	− 10.4**	-4.5
	(2.7)	(2.9)	(1.3)	(3.3)	(2.6)
Non-American born parent	− 1.0	− 1.7*	− 0.6	− 1.8*	− 0.6
	(0.6)	(0.7)	(0.3)	(0.9)	(0.6)
Household (1910)					
Family size	− 1.1***	− 1.1***	− 0.3***	− 0.8***	− 0.8***
	(0.1)	(0.1)	(0.1)	(0.2)	(0.1)
Families in household	0.6	0.6	0.2	0.6	0.2
	(0.3)	(0.4)	(0.2)	(0.4)	(0.3)
Non-family household	3.7	4.8	1.1	3.2	− 0.4
	(4.2)	(4.8)	(2.1)	(6.1)	(3.1)
Urban household	3.6***	3.2**	1.2**	1.8	2.8**
	(1.0)	(1.0)	(0.4)	(1.0)	(0.9)
Farm household	− 1.2	− 2.9**	− 1.1**	− 3.3**	− 0.3
	(0.9)	(1.1)	(0.4)	(1.1)	(0.8)
Paying mortgage	− 0.3	0.4	− 0.0	0.5	− 0.6
	(0.5)	(0.6)	(0.2)	(0.7)	(0.5)
Parents (1910)					
Mother: Age	0.3***	0.4***	0.1***	0.2***	0.2**
	(0.1)	(0.1)	(0.0)	(0.1)	(0.1)
Mother: Non-American born parent	0.4	− 0.3	0.2	− 0.1	0.2
	(0.5)	(0.6)	(0.2)	(0.6)	(0.5)
Father: Age	− 0.1	− 0.1	− 0.0	− 0.1	− 0.0
	(0.1)	(0.1)	(0.0)	(0.1)	(0.0)
Father: Non-American born parent	0.6	− 0.4	− 0.1	− 1.0	0.2
	(0.6)	(0.8)	(0.3)	(0.8)	(0.6)
Father: Education score	0.2***	0.2***	0.1***	0.1***	0.3***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Earnings score	0.1***	0.1***	0.0***	0.1***	0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Unemployed	1.4	1.6	0.5	1.4	1.2
	(1.1)	(1.2)	(0.5)	(1.4)	(1.0)
Controls					
Industry father	✓	✓	✓	✓	✓

Table 6 (continued)

	Composite		Single dimension		
	(1)	(2)	(3)	(4)	(5)
	SEI	NPBOSS	OCC.INC	OCC.ER	OCC.ED
County characteristics	✓	✓	✓	✓	✓
State indicator	✓	✓	✓	✓	✓
R^2	0.12	0.12	0.10	0.09	0.08
N	9042	9162	9184	9162	9139

Notes: (i) stars indicate significance according to $*p < 0.05$, $**p < 0.01$, $***p < 0.001$; (ii) this table shows the results of the Ordinary Least Squares regression shown in Eq. (3) estimating the effect of wildfire smoke exposure in early childhood on later-life socioeconomic status conditional on controls; (iii) the data are obtained from the Integrated Public Use Microdata Series full-count censuses 1910 and 1940; (iv) the composite socioeconomic status measure SEI stands for the Duncan Socioeconomic Index, and NPBOSS is the Nam–Powers–Boyd Occupational Status Score; (v) the single dimension socioeconomic status measures are occupational income (OCC.INC), earnings (OCC.ER), and education (OCC.ED)

status. Regarding parental characteristics, our results suggest that an increase in the age of the mother, as well as the father's Occupational Earnings Score, is positively associated with better ranking on later-life occupational standings. Reassuringly, these results are consistent with those in the traditional labour economics literature.

Table 6 presents the results on men linked in 1940 revealing negative effects of early-childhood smoke exposure on a majority of occupational standings indicators. Notably, the single dimension indicator Occupational Income is significant indicating a decrease of 90 US$ in annual income comparing men impacted by smoke to those not affected. This represents a 3.6% reduction compared to the 1940 annual average occupational income of 2,490 US$ (Table 2).[26] While in 1930 the results reveal a negative effect of smoke exposure on the composite measures with no clear indication of which single dimension is driving the effect, the findings in 1940 suggest that men who were exposed to wildfire smoke in early childhood experience a negative income effect in 1940.

More generally, one should note that the adverse effect of early-childhood wildfire exposure on later-life socioeconomic status indicators found 20–30 years after the wildfire may have potentially arisen through health effects that affected both physiological and cognitive aspects of development. From a physiological perspective, the detrimental effects of wildfire smoke exposure are especially severe during early childhood, a critical period when the body is still developing. For instance, the presence of incomplete barriers in young children allows a greater proportion

[26] The IPUMS 1940 census data are the first one to provide reported income (as oppose to income derived from an individual's occupation). We find that there is a negative insignificant effect of smoke exposure on reported income in 1940 but do not report this in the tables for the sake of consistency across the censuses.

of harmful particles to penetrate deeper into their lungs compared to adults, leading to an increased risk of respiratory morbidity, including conditions such as asthma, bronchitis, and other chronic pulmonary diseases (Bennett et al. 2007). These negative health impacts may result in missed school days or impaired learning, ultimately affecting children's educational outcomes and overall quality of life. Besides the potential effects on respiratory conditions, early-life exposure to smoke pollution may impact physiological growth, which in turn is found to be associated with adulthood income as shown in Tan-Soo and Pattanayak (2019). The authors build on the findings of Rosales-Rueda and Triyana (2019), who reported an association between early-childhood smoke exposure and reduced stature by the age of 17 and highlight the link between adult height and income, proposing that a decrease in height due to exposure to agricultural fires could lead to an average reduction of approximately 4% in monthly income during adulthood.

Considering the limited research on the impact of wildfire smoke exposure on cognitive and neuropsychological development in children, especially those under the age of five, it is beneficial to also consider evidence from exposure to $PM_{2.5}$ from other sources. In the literature review on air pollution and neuropsychological development in children by Suades-González et al. (2015), the authors report that there is inadequate or insufficient evidence on the association between $PM_{2.5}$ and cognitive ad psycho-motor development. However, the authors suggest a positive link between post-natal $PM_{2.5}$ exposure and autism spectrum disorder (ASD). For example, Talbott (2015) conduct a population-based case–control study in Pennsylvania and find a significant association of $PM_{2.5}$ exposure at the age of two and childhood ASD. This evidence aligns with the findings of Brockmeyer and D'Angiulli (2016), who highlight that in early childhood the blood–brain barriers are not fully developed, allowing particulate matter to cause significant neuroinflammation and subsequent cell loss in the central nervous system. These neurological impacts may lead to cognitive deficits, which can challenge children's educational attainment, thereby limiting their opportunities for higher-paying jobs and contributing to lower socioeconomic status outcomes in adulthood. Furthermore, these cognitive challenges may discourage parents from sending their children to school, further exacerbating the long-term impact on their socioeconomic prospects. Note that while many studies assess prenatal exposure to air pollution, evidence for young children is scarce (Suades-González et al. 2015).

It is noteworthy that the coefficient of the burned area indicator is significantly positive for the Duncan Socioeconomic Index in 1930.[27] Similarly, when excluding the fire-affected areas from the estimation completely, all of the point estimates but occupational education are negative and significant and larger in magnitude as shown in Table 17 and Table 18 for 1930 and 1940, respectively.[28] This indicates that the fire-affected regions may have experienced a positive shock (along the lines

[27] Nevertheless one should keep in mind that, even after controlling for our rich set of covariates, the validity of the causal interpretation of having been in a fire burnt county is likely weaker than for our smoke exposure proxy.

[28] As this estimation leaves only four counties categorised as treated, and the sample size is reduced substantially, we exercise caution by abstaining from interpreting the magnitude of the coefficients.

of "creative destruction") of burned area counteracting the purely negative shock of being only smoke-affected. One possible mechanism behind this finding could be the fact that the timber industry saw a surge in significance following the Big Burn. This increase was not only due to its contributions to fire prevention via logging, but also because it received enhanced support from the USFS – an evolution detailed by subsequent developments. After the Great Fire of 1910, President William Taft replaced USFS Chief Forester Gifford Pinchot, a staunch conservationist, due to a disagreement on land management policies, leading to a shift in the USFS's stance towards timber industry-friendly policies (Egan 2009, p. 275). Pinchot's successor, Henry Graves emphasised the timber industry's enhanced role, stating: "The public forests are being protected from fire, the timber is used as it is called for by economic conditions, and the cutting is conducted by such methods as leave the land in favourable condition for the next crop of timber".[29] Additionally, Graves founded the Forest Products Laboratory in Madison, Wisconsin, to boost lumber production, wood preservation, and reduce logging waste, likely enhancing timber productivity. Concurrently, William Greeley, a regional forester during the Great Fire, rose to a high administrative USFS role, advocating the logging industry's involvement as a fire prevention co-manager. By 1920, he ascended to USFS chief and, eight years later, joined a timber syndicate as an executive (Egan 2009, p. 270, 271).

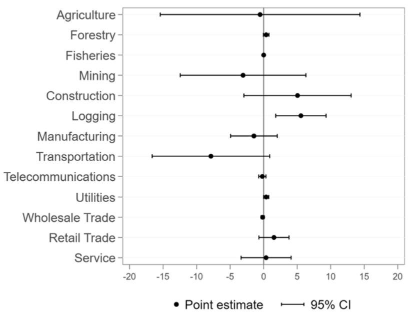
To roughly ascertain whether the alleged growth in the timber industry in counties that experienced fires is also apparent in the data, we conduct a two-way fixed effects Difference-in-Differences estimation of the shift in county-level industry shares from immediately before the fire (1910) to 1920 using the census data. More specifically, we estimated the change in the labour market's industry shares by comparing fire-affected counties with (i) counties affected solely by smoke, as shown in Fig. 5a, and (ii) all remaining counties within the study area, depicted in Fig. 5b. As can be seen, the timber industry boom following the Great Fire of 1910 occurred mainly in the counties which had burned areas compared either to the counties that were only smoke-affected or to all other counties.

5.3 Robustness checks

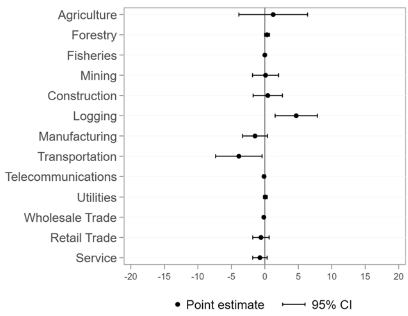
The baseline estimates, the specification dropping the fire-affected counties as well as a number of robustness checks for coefficients of interest for the short-term and later-life analyses are shown in Fig. 6. Note that not all robustness checks are suitable for presentation in that format and certain later-life estimations encounter substantial constraints due to a large reduction in sample size.

First, we conduct a number of permutation tests in the spirit of Fisher (1937). More specifically, we randomise which of the 14 out of the 70 counties are smoke-affected, and then also randomise in which 10 of these 14 are also fire-affected. Subsequently, we run Eqs. (2) and (3) performing 1000 iterations and plot the distribution of the corresponding t-statistic. The p value is derived by the rank of the t-statistic of the main estimation. For the short-term excess mortality result in the

[29] The quote is obtained from the Forest History Society under https://foresthistory.org/research-explore/us-forest-service-history/people/chiefs/henry-s-graves-1871-1951.



(a) Comparison counties: smoke-affected

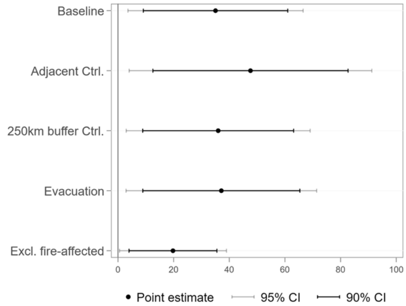


(b) Comparison counties: all

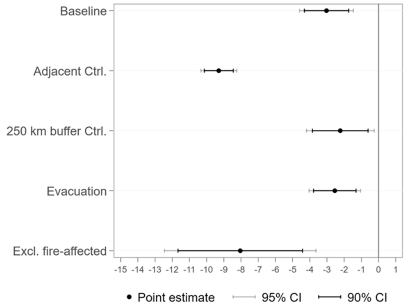
Fig. 5 Two-way fixed effects Difference-in-Differences regression of industry share on counties comprising burned areas. *Notes:* (i) the plots show the point estimates and 95% confidence intervals estimating a Difference-in-Difference regression with county and year fixed effects of industry share on burned area as an indicator variable; (ii) standard errors are clustered at the county level; (iii) the comparison group in **a** includes the counties that were solely smoke-exposed and **b** displays the estimation using all counties in the study area as comparison group; (iv) the industry data are derived from the Integrated Public Use Microdata Series USA 1910 and 1920 censuses and include only individuals in the labour force

week of the wildfire, this permutation test indicates a *p* value of 0.008 (1–992/1000) as shown in Fig. 11a in Appendix B which demonstrates that the result is unlikely to be driven by chance. For the later-life socioeconomic outcomes, the similar test is performed for the estimates that are significant in 1930, i.e. the Duncan Socioeconomic Index and the Nam–Powers–Boyd Occupational Status Score as shown and for Occupational Income in 1940. While the actual t-statistic of the Duncan Socioeconomic Index (Fig. 11b in Appendix B) is highly significant with a *p* value of 0.015 (15/1000), the *p* value of the Nam–Powers–Boyd Occupational Status Score (Fig. 11c in Appendix B) is 0.094 (94/1000), and thus significant at the 10% level. Regarding the 1940, Occupational Income in this permutation exercise has a *p* value of 0.08 (80/1000) and is significant at the 10% level (Fig. 11d in Appendix B).

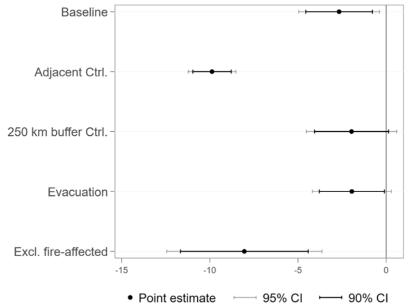
Given excess deaths can be sensitive to analytical choices we estimate the excess mortality rate in Eq. (2) implementing a multiverse analysis as suggested in Levitt et al. (2023). First, in addition to our baseline reference period of the prior 5 years, we use reference periods ranging from 2 to 4 years shown in Fig. 12 in Appendix B. Similar to our main estimation, the excess mortality rate of children under the age of five ranges from 34.5 to 36.3 per 100,000. Second, since more early years may not be as relevant as later years, we combine the reference years with (i) weights decreasing linearly by 5% (i.e. 100% weight for 1909, 95% weight for 1908, etc.), (ii) decreasing linearly by 10%, and (iii) weights decreasing by half for each year before 1910. As shown in Fig. 13 the point estimate indicating the excess mortality rate of children is estimated at 32 to 34.7 deaths per 100,000, and thus very close to our main estimation. Third, we use a Quasi-Poisson model accounting for the overdisperison in the mortality data and including a time trend to estimate the excess mortality rate. The results shown in Fig. 14 indicate a slightly lower excess mortality rate of 31.3 children per 100,000 under the age of five with a *p* value of 0.06. Note, that this model may not be most suitable in our context given that we already use a



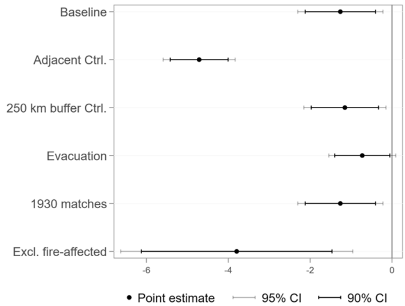
(a) Number of excess deaths in the fire week per 100,000



(b) Duncan Socioeconomic Index 1930



(c) Nam-Powers-Boyd Occupational Status Score 1930



(d) Occupational Income 1940

Fig. 6 Model specifications and robustness checks overview for short-term and long-term estimations. *Notes:* (i) **a** presents the point estimates, the 95%, and 90% confidence intervals of the short-term analysis delineated in Eq. (2); **b**, **c** and **d** show the point estimates, the 95%, and 90% confidence intervals of the later-life outcomes estimation as specified in Eq. (3); (ii) "Baseline" refers to the primary analysis; "Adjacent Ctrl." encompasses only the counties bordering the treatment counties included in the comparison group; "250 km buffer Ctrl." includes solely the counties within a 250 km buffer of the smoke plume as comparison counties; "Evacuation" denotes the exclusion of the two counties Spokane and Missoula; "Excl. fire-affected" keeps the only smoke-affected counties as the treatment group; "1930 matches" retains individuals linked from 1930 in the 1940 dataset without adding new matches

county-level age-stratified sample accounting for seasonality by estimating weekly deaths, while Poisson models may be more effective for yearly all-age mortality data.

One might be worried that too many comparison counties are included in our estimation that are peripheral to the source. To investigate this we run two different estimations reducing the number of comparison counties (i) to those directly adjacent to the treatment counties, i.e. sharing a boundary (16 comparison counties), and (ii) using a calculated buffer area around the modelled smoke plume of 250 km, and include all the counties in the comparison group that are within that buffer (40 comparison counties). The corresponding maps are shown in Fig. 15 in Appendix B. The estimations of Eq. (2) indicate higher weekly excess mortality rate coefficients of 47.6 and 36 per 100,000 for the estimation with 16 and 40 adjacent comparison counties, respectively. Both the Difference-in-Differences plots (Fig. 16) and the regression tables (Tables 19 and 20) are shown in Appendix B. As for the long-term impact on later-life socioeconomic outcomes, all the 5 SES indicators are significantly negative in 1930 for the estimation with including only the 16 adjacent counties and the Duncan Socioeconomic Index is significant in the 250 km buffer estimation. Occupational Income is negative significant for both estimations in 1940. The long-run coefficients of interest are included in Fig. 6. Nevertheless, given the substantial decrease in sample size, we exercise caution in interpreting the coefficients for this robustness check regarding the later-life outcomes.

To address concerns over the potential impacts of evacuations we assess whether our results are robust to the exclusion of Spokane County, Washington and Missoula County, Montana, which are the two destination counties of train evacuations from the immediate burn zone (Krainz 2012). Note, that very few residents were evacuated and that evacuations were administered rather late when the population arguably was already smoke-exposed.[30] We find that the excess mortality rate estimate for the fire week is robust to dropping these counties from the comparison group with a coefficient of 37.1 per 100,000, which is slightly larger than our baseline result of 35 per 100,000 (see Fig. 17 for the Difference-in-Differences plot and Table 21 for the regression table). In terms of the long-term estimation we still find significant negative effects for the Duncan Socioeconomic Index with a reduction of 2.5 points in 1930 and negative effects for the Nam–Powers–Boyd Occupational Status Score at the 10% level (Table 22 in Appendix B). Furthermore, Occupational Income in 1940 remains significant at the 10% level (Table 23 in Appendix B). Related to the topic of evacuation are concerns regarding migration. Yet, the duration of evacuee displacement, was notably brief. Krainz (2012) describe that the refugees returned within days. For instance only 175 women and children from an initial group of 1,200 remained in Missoula after just one week. This short duration of displacement suggests that evacuation did not lead to long-term resettlement. Nonetheless, it does not preclude the possibility of later moves unrelated to the immediate evacuation. However, as indicated in Sect. 4.1 we find no evidence that smoke exposure in 1910 influenced the likelihood of individuals relocating to different counties in any of the subsequent census years of 1920, 1930, or 1940.

Finally, we re-estimate Eq. (3) for 1940 by exclusively retaining the links established in 1930, thereby maintaining a consistent cohort of men throughout the study period as opposed to using the newly established links. The findings visually illustrated in Fig. 6d denoted as “1930 matches” and as a regression output in Table 24 in Appendix B suggest that early-life smoke exposure leads to an even larger negative effect on Occupational Income of 130 US$ compared to the baseline estimations, which corresponds to a decrease of 5.2% relative to the 1940 average annual Occupational Income (Table 2). Moreover, restricting the analysis to following the 1930 cohort also shows a decrease in 1940 Occupational Earnings of 3.6, which translates to a decrease of 7.3% compared to the 1940 average (Table 2). Thus, we show that the negative effect found in 1930 persists in 1940 for the same set of individuals.

[30] This assumption is corroborated by newspaper articles reporting on evacuations. Example quotes can be requested from the authors.

6 Conclusions

In this study, we assessed the smoke-induced short-term mortality and long-term socioeconomic impacts of the Great Fire of 1910 for smoke-exposed children under the age of five. To this end, we used historical burn perimeters within a wildfire smoke emission and dispersion model to proxy smoke exposure and combined this with mortality records and linked full-count census data from 1900 to 1940. Our econometric estimations suggest a short-term effect on the excess mortality rate for children residing in smoke-affected regions in the week of the wildfire. Furthermore, we find evidence that boys who were under the age of five at the time of the Great Fire of 1910 ranked lower on some socioeconomic status indices in 1930 and 1940 if they resided in smoke-affected counties compared to boys who did not.

Our paper provides novel evidence both in the economic assessment of smoke exposure during wildfires in terms of short-term health effects during childhood and later-life socioeconomic outcomes in a historic context where avoidance behaviour was likely relatively limited.

More generally, this study arguably contributes to a deeper understanding of the implications of major wildfire events for public health and human capital formation in any context where avoidance may be relatively limited. While awareness of the adverse health effects of wildfire smoke exposure has certainly increased substantially since 1910, avoidance options are unequal between countries, communities, and neighbourhoods (Grant and Runkle 2022).

A. Appendix

See Figs. 7, 8 and Table 7.

A.1 County boundary changes

To account for boundary changes from 1900 to 1910 we use (i) the Big Ten Academic Alliance (BTAA) Geoportal which provides historical shape files for the years 1900 and 1910[31] and (ii) the Atlas of Historical County Boundaries which documents county-level boundary changes for all US states.[32] For counties affected by boundary changes between 1900 and 1910 we approximate the population in 1900 based on the county boundaries of 1910 taking account of the week of the enforcement of the change. For example, on 21st February 1907 (International Organization for Standardization (ISO) week 8), county Kootenai was split into counties Kootenai and Bonner. Thus, a theoretical 1900 Bonner County would have consisted of 61% of 1910 county Kootenai and the 1900 Kootenai County would only be 39% of the 1910 Kootenai County. Hence, for all weeks before the boundary change, the 1900

[31] The historical shape files are available under https://geo.btaa.org/ e.g. https://geo.btaa.org/catalog/harvard-nhgis-pop1910 for the year 1910.

[32] For more detail see https://digital.newberry.org/ahcb/project.html.

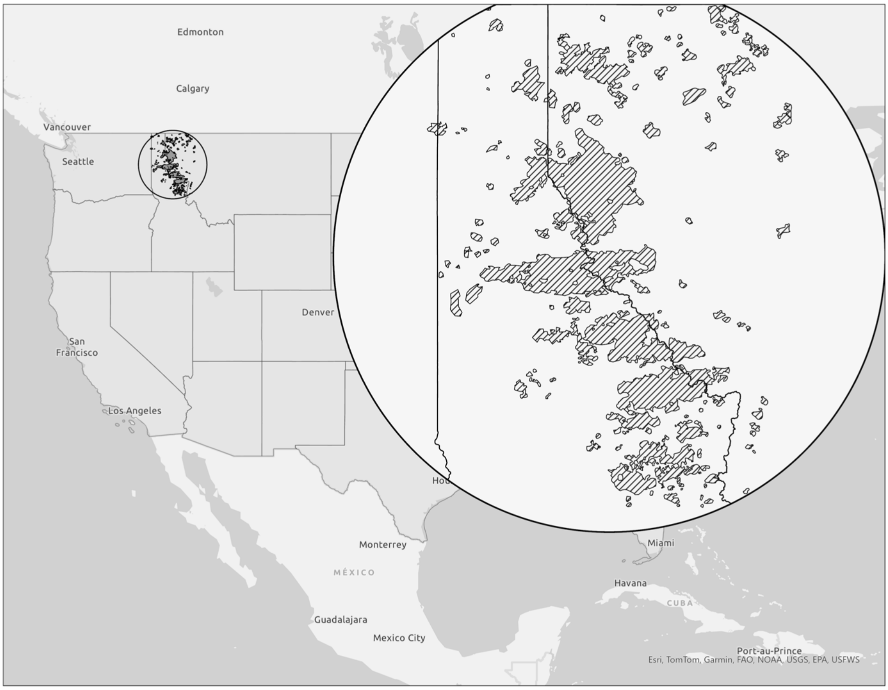


Fig. 7 Location of the Great Fire of 1910 within the US, which burnt in the states of Idaho, Montana, and Washington. *Note:* The above map shows the extent of burned area and is created in ArcGIS using the georeferrenced burn scar polygons

population of county Kootenai and a hypothetical 1900 Bonner County is adjusted accordingly. The summary of the calculations of these boundary changes is shown in Table 8.

See Tables 9, 10 and Fig. 9.

See Fig. 9 and Tables 11, 12, 13 and 14.

See Fig. 10 and Tables 15, 16, 17 and 18.

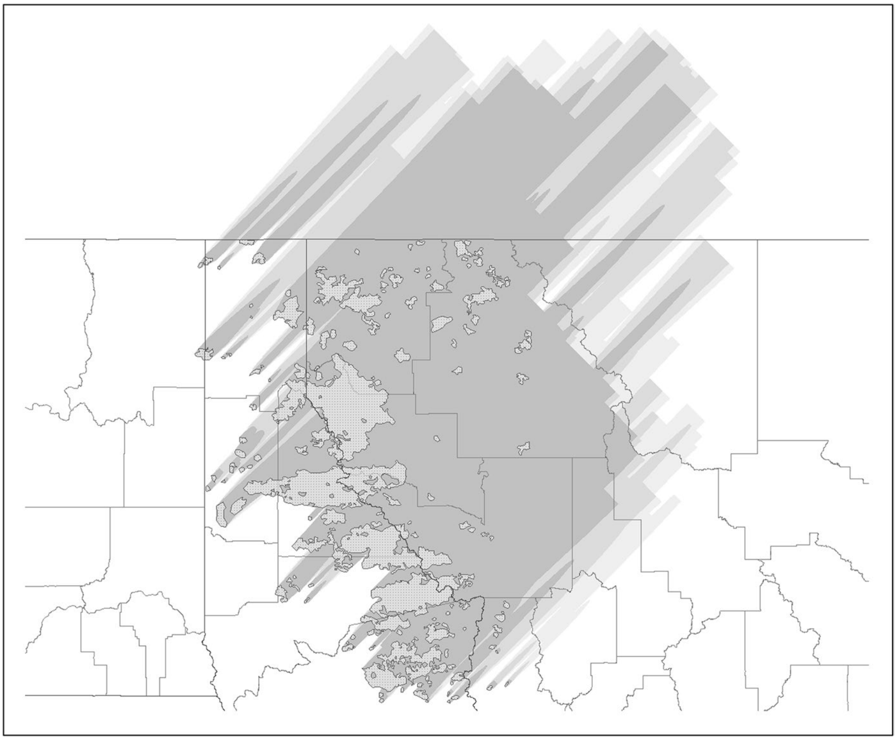


Fig. 8 Modelled smoke plumes generated using BlueSky's default settings. *Notes:* (i) the smoke plumes shown in in this image are modelled using BlueSky Playground version 3.5.1; (ii) in contrast with Fig. 1b, the modelling process runs with the framework default settings, i.e. the moisture level is "Dry", the stability class is "Near neutral", and the surface relative humidity value is 25%; (iii) the grey dotted area shows the burn perimeters. The darkest grey-shaded area indicates the area where the hourly peak $PM_{2.5}$ pollution is hazardous ($PM_{2.5}$ > 526 μg/m^3). The medium grey area shows unhealthy hourly peak $PM_{2.5}$ pollution of the values ($PM_{2.5}$ > 130 μg/m^3) and the lightest grey scale denotes moderate hourly peak pollution ($PM_{2.5}$ > 38 μg/m^3)

Table 7 Sources in Egan (2009) for model input specifications in the BlueSky smoke emission and dispersion framework

Model input	Selected value	Source(s) by Egan (2009)
Emission inputs		
Moisture level	Very dry	"People had endured a summer without moisture, the driest in a generation... (Egan 2009, p. 227)"
		"[19th of August] when Emma Pulaski walked outside of town [Wallace] to greet her husband, and all the vegetation crunched underfoot, brown and crisp to the touch, she knew that everything and everybody in this pocket of people in the mountains had been reduced to fuel (Egan 2009, p. 143)"
		"the wind sucked moisture from the forest, blowing hard in the afternoons, leaving all that standing timber as if it was just hung out to dry (Egan 2009, p. 108)"
Number of days	1	"For forty-eight hours, no one knew whether they would see another day or recognize their homes again (Egan 2009, p. 227)"
Combustion phase	Flaming	"... single wall of yellow and orange... and it burned at the crowns, the highest tips of the trees exploding into the air, flying off to light the crowns of other tall trees (Egan 2009, p. 115)"
Dispersion inputs		
Stability class	Moderately unstable	"... the Palouse is one of those curious places in the West where a weather system can form benign and transform into something ferocious long after it has left the cradle of its creation.... the air over the Palouse can be volatile, or violent. So it was on the Saturday afternoon of August 20th, when atmospheric conditions gave birth to a Palouser that lifted the red dirt of the hills and slammed into the forests – not as a gust or an episodic blow, but as a battering ram of forced air (Egan 2009, p. 154)"
		"The chain reaction of a wildfire had begun. Heated plant matter released hydrogen and carbon while drawing in oxygen, and the whole of it was on the run, a weather system of its own (Egan 2009, p. 155)"
		"Looking up, he saw towering columns of black smoke above the treeline, and then – the strangest of things – the columns themselves would explode into flame, sending off fire hundreds of feet above the treetops, like the towering spigot of a refinery. What he saw – known as fire whirls – can reach temperatures of 2000°, with a downdraft in the center and violent updrafts on the outer rings (Egan 2009, p. 159)"
Wind direction	225°	"Everyone knew about Palousers, the warm winds from the southwest (Egan 2009, p. 4)".
		"... a lumbering advance of flame pushed along by the prevailing winds from the south and west (Egan 2009, p. 145)"
Surface relative humidity	10%	"He [Billy Greeley, Forest Service's Region 1 forester] had never seen the woods so ready to explode. Low humidity had dropped to the level of the Mojave Desert (Egan 2009, p. 114)"

Notes: (i) the above table explains the data sources upon which our inputs for the BlueSky modelling framework are selected; (ii) BlueSky Playground version 3.5.1 is employed to model the smoke emission and dispersion

Table 8 County boundary changes from 1900 to 1910

FIPS	County 1910	State	Matching Area 1900	Comment
16017	Bonner	ID	61% Kootenai	21 Feb 1907 (W08): Kootenai split in Kootenai and Bonner
16031	Cassia	ID	57% Cassia	21 Feb 1907 (W08): Cassia split in Cassia and Twin Falls
16055	Kootenai	ID	39% Kootenai	21 Feb 1907 (W08): Kootenai split in Kootenai and Bonner
16069	Nez Perce	ID	100% Nez Perce & 53% Shoshone	1904: Nez Perce gained from Shoshone
16079	Shoshone	ID	47% Shoshone	1904: Shoshone lost to Nez Perce
16083	Twin Falls	ID	43% Cassia	21 Feb 1907 (W08): Cassia split in Cassia and Twin Falls
30017	Custer	MT	58% Custer	1901: Custer split in Custer and Rosebud
30023	Deer Lodge	MT	14% Deer Lodge & 32% Silver Bow	1901: Deer Lodge lost to Powell 1903 Deer Lodge gained from Silver Bow
30029	Flathead	MT	64% Flathead	1 Jul 1909 (W26): Flathead split in Flathead and Lincoln
30053	Lincoln	MT	36% Flathead	1 Jul 1909 (W26): Flathead split in Flathead and Lincoln
30063	Missoula	MT	56% Missoula	1 Mar 1906 (W09): Missoula split in Missoula and Sanders
30077	Powell	MT	84% Deer Lodge	1901: Powell created from Deer Lodge *
30087	Rosebud	MT	42% Custer	1901: Custer split in Custer and Rosebud
30089	Sanders	MT	44% Missoula	1 Mar 1906 (W09): Missoula split in Missoula and Sanders
30093	Silver Bow	MT	68% Silver Bow	1903: Silver Bow lost to Deer Lodge
53005	Benton	WA	24% Yakima & 17% Klikatat	8 Mar 1905 (W10): Benton created from Klikatat and Yakima
53017	Douglas	WA	40% Douglas	24 Feb 1909 (W08): Douglas split in Douglas and Grant
53025	Grant	WA	60% Douglas	24 Feb 1909 (W08)16: Douglas split in Douglas and Grant
53039	Klikatat	WA	80% Klikatat	8 Mar 1905 (W10): Part of Klikatat lost to creation of Benton
53077	Yakima	WA	76% Yakima	8 Mar 1905 (W10): Part of Yakima lost to creation of Benton

*0.3% of the area was used from Lewis and Clark but we do not consider attributions lower than 1%; (i) the exact taking effect date of the boundary change is only indicated if it affects the mortality data, i.e. changes from 1905 to 1909; (ii) in the "Comment" column the ISO week is indicated in parentheses; (iii) the shape files of the county boundaries in the year 1900 and 1910 are obtained from https://geo.btaa.org and the exact information on the boundary changes is from https://digital.newberry.org/ahcb/project.html; (iv) FIPS denotes the abbreviation for Federal Information Processing Standard and is a 5-digit county identification code

Table 9 Ancestry data sources for county-level deaths from 1905 to 1910

General
USA, Find a Grave Index, 1600 s-Current
USA, Newspapers.com Obituary Index, 1800 s-current
USA, Presbyterian Church Records, 1701–1970
Montana
Beaverhead County, Montana, USA, County Records, 1862–2009
Missoula and Ravalli County, Montana Cemeteries
Montana, USA, County Births and Deaths, 1830–2011
Montana, USA, State Deaths, 1907–2018
Web: Gallatin County, Montana, USA, Death Index, 1856–2014
Idaho
Idaho, USA, County Birth and Death Records, 1863–1970
Idaho, USA, Death Index, 1890–1964
Idaho, USA, Death Records, 1890–1969
Idaho, USA, Select Deaths and Burials, 1907–1965
Salt Lake City, Utah, USA, Cemetery Records, 1848–1992
Salt Lake County, Utah, USA, Death Records, 1908–1949
Washington
Washington, USA, Death Index, 1940–2017
Washington, USA, Death Records, 1883–1960
Washington, USA, Select Death Index, 1907–1960

Note: The source for this table is the genealogy company Ancestry.com

Table 10 The individual steps of the data cleaning process of the mortality data

1. Cleaning date of death
Remove non-numeric characters, e.g. “Ab”, “abt”, “about”, “unknown date”*
Drop observations if the date of death is longer than 11 characters
Drop observations with year range, e.g. 1900–1913
Drop observations only including the month and year of death but not the day
2. Cleaning names and gender
Drop observations if name includes “Void”
Remove special characters from names, e.g. numbers, “Sgt”, “1lt”*
Drop observations if the name includes characters such as “?,” “...,” “#,” “$”
Identify gender based on indications in names, e.g. “Mrs.,” “Mr.,” “Daughter,” “Son,” “Girl,” “Boy”*
If gender is missing applying historical gender prediction function using R-package `gender` defining year range as 1800–1915 (Mullen 2021)
3. Cleaning date of birth and age at death
Remove characters from date of birth, e.g. “ab”, “about”, “unknown”, “not obtainable”, “Don’t know about”*
Replace the birth date with NA if no characters are left
Add 1st of respective month if only birth month and year are given in the date of birth

Table 10 (continued)

Drop observations if the birth year is missing
Adjust all two-digit years to four-digit years
Drop observations with age zero and negative age**
Infer age from date of birth and date of death if age is missing
4. Removing duplicates
Drop duplicate observations if several observations are identical on the variables name, gender, date of death, state, and county FIPS
Drop duplicate observations if several observations are identical on the variables date of birth, date of death, age at death, and county FIPS
Drop duplicate observations if several observations are identical on the variables date of birth, name, age at death, and county FIPS
Drop duplicate observations if several observations are identical on the variables gender, name, birth year, death year, and county FIPS

*These steps also include an extensive number of typos and misspellings; **We drop the observations with age zero because with the use of different data sources it would not be possible to detect duplicates as some records denote a death before the age of one as "Stillborn", zero, or an indication in weeks or months. *Notes:* (i) The above table shows the specific steps that are undertaken to clean the digitised mortality data obtained from Ancestry.com; (ii) FIPS denotes the abbreviation for Federal Information Processing Standard and is a 5-digit county identification code

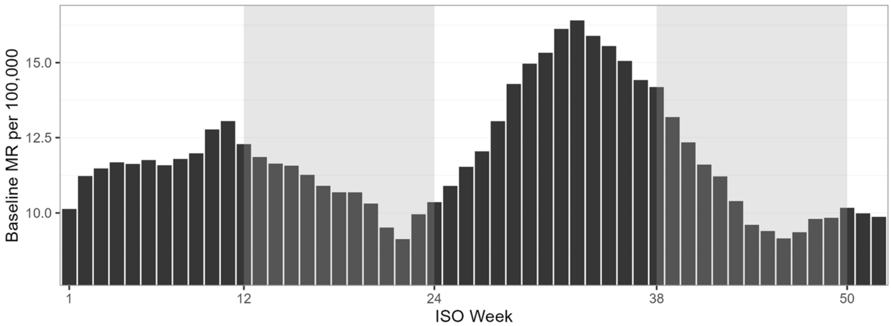


Fig. 9 Seasonal course of the baseline mortality rates for children under the age of five in the study area. *Notes:* (i) Baseline MR indicates the baseline mortality rate for children under five and ISO stands for International Organization for Standardization; (ii) the mortality data are obtained from the genealogy company Ancestry.com and the population data are retrieved from the 1900 and 1910 full-count US censuses provided by the Integrated Public Use Microdata Series (IPUMS) USA; (iii) the grey-shaded areas show the meteorological spring (ISO weeks 12–24) and autumn (ISO weeks 38–50)

Table 11 Descriptive statistics of the control variables at the county level for the 70 sample counties in 1910

	Min	Mean	SD	Median	Max
Average age	22.0	28.2	2.1	28.5	32.2
Average Socioeconomic indicator	15.4	20.4	2.6	19.9	30.1
Average family size	2.9	4.0	0.7	3.8	6.2
Number of families in household	1.2	2.8	0.9	2.8	5.3
Non-family household (%)	2.3	9.3	4.6	8.4	22.9
Place > 1000 habitants (%)	0.0	11.7	13.0	10.5	48.7
Farm household (%)	2.3	39.5	15.1	42.1	73.4
Rented property (%)	15.4	33.0	9.9	33.6	55.0
Paying mortgage (%)	4.4	24.8	12.1	24.3	56.6
Multigenerational household (%)	68.7	81.3	5.6	81.9	92.6
Non-white (%)	0.0	3.3	4.8	1.7	30.0
Non-American parent (%)	21.3	45.1	11.9	45.0	76.8
In school (%)	11.6	19.2	4.8	18.1	32.8
Unable to read and write (%)	0.5	4.3	3.7	3.3	20.4
Unemployed (%)	1.1	6.1	3.5	5.5	17.3
In labour force (%)	3.3	8.2	3.7	7.3	26.2
Ind. Agriculture, forestry & fishing (%)	4.2	48.3	18.1	52.4	78.4
Ind. Mining (%)	0.0	7.0	11.1	1.8	53.4
Ind. Construction (%)	1.1	6.6	4.0	5.4	23.5
Ind. Manufacturing (%)	1.5	7.5	8.5	4.5	50.1
Ind. Transportation, communication & utilities (%)	2.5	13.4	8.3	11.6	37.4
Ind. Wholesale and retail trade (%)	3.8	7.7	2.5	7.4	15.3
Ind. Finance, insurance, and real estate (%)	0.3	1.3	0.7	1.2	4.3
Ind. Business and repair services (%)	0.4	1.1	0.4	1.1	2.3
Ind. Personal services (%)	1.3	3.2	1.2	2.9	8.4
Ind. Entertainment and related services (%)	0.0	0.3	0.2	0.3	1.0
Ind. Professional and related services (%)	0.8	2.2	0.7	2.1	4.1
Ind. Public administration (%)	0.3	1.4	1.3	1.0	9.0

Notes: (i) *SD* standard deviation, *Ind.* industry; (ii) the variables are obtained from the Integrated Public Use Microdata Series USA 1910 full-count individual and household censuses

Table 12 Descriptive statistics of the 1910 variables including individual, household, and parental characteristics of 7801 boys linked in 1930

	Min	Mean	SD	Median	Max
Individual					
Non-white (%)	0.0	0.7	8.2	0	100
Non-American born parent (%)	0.0	28.0	44.9	0	100
Household (1910)					
Family size	3.0	6.0	2.2	6	19
Families in household	1.0	1.3	0.7	1	8
Non-family household (%)	0.0	0.3	5.9	0	100
Urban household (%)	0.0	17.1	37.6	0	100
Farm household (%)	0.0	59.2	49.1	100	100
Paying mortgage (%)	0.0	35.6	47.9	0	100
Mother (1910)					
Age	16.0	31.4	6.8	31	66
Non-American born parent (%)	0.0	46.5	49.9	0	100
Father (1910)					
Age	18.0	37.3	8.0	37	70
Non-American born parent (%)	0.0	48.5	50.0	0	100
Education score	0.0	8.6	12.7	5	94
Earnings score	1.4	29.9	29.9	10	100
Unemployed (%)	0.0	5.1	21.9	0	100
Ind. Agriculture, forestry & fishing (%)	0.0	63.6	48.1	100	100
Ind. Mining (%)	0.0	3.7	18.8	0	100
Ind. Construction (%)	0.0	6.4	24.5	0	100
Ind. Manufacturing (%)	0.0	5.7	23.1	0	100
Ind. Transportation, communication & utilities (%)	0.0	5.1	22.0	0	100
Ind. Wholesale and retail trade (%)	0.0	8.0	27.1	0	100
Ind. Finance, insurance, and real estate (%)	0.0	1.6	12.7	0	100
Ind. Business and repair services (%)	0.0	1.1	10.3	0	100
Ind. Personal services (%)	0.0	1.7	12.9	0	100
Ind. Entertainment and related services (%)	0.0	0.1	3.8	0	100
Ind. Professional and related services (%)	0.0	1.6	12.7	0	100
Ind. Public administration (%)	0.0	1.3	11.4	0	100

(i) note that all the variables in percentages (%) are indicator variables and therefore, a median of 100 means that the majority of observations are of the value 1, and 0 means that the majority of observations are of the value 0; similarly, a maximum value of 100 for indicator variables in percentages (%) denotes that the maximum value is 1; (ii) *SD* standard deviation, *Ind.* industry; (iii) the variables are obtained from the Integrated Public Use Microdata Series USA 1910 full-count individual and household censuses

Table 13 Probit model assessing the likelihood of an individual living in a different county in 1920, 1930, and 1940 relative to the county of residence in 1910

	Individual moved counties (1/0)		
	1920	1930	1940
Smoke	0.2	− 0.2	− 0.2
	(0.4)	(0.4)	(0.3)
Burned area	− 0.0	0.2	0.2
	(0.4)	(0.4)	(0.4)
Controls			
Individual	✓	✓	✓
Household	✓	✓	✓
Parents	✓	✓	✓
County	✓	✓	✓
State	✓	✓	✓
N	12,362	7801	5550

Notes: (i) the table shows the probit model estimating the likelihood of an individual moving counties between 1910 and 1920, 1930, or 1940, respectively, conditional on being smoke-affected. The model controls for all control variables used in Eq. (3); (ii) individual level variables are non-white and non-American born parent; household level variables include family size, number of families in a household, family household, urban household, farm household, and paying mortgage; parental controls are age and non-American born parent of the mother and age, non-American born parent, education score, earnings score, unemployed, and industry of the father; county-level variables are average age, average socioeconomic indicator, average family size, number of families in household, percentage non-family household, percentage places larger than 1000 habitants, percentage farm households, percentage rented property, percentage paying mortgage, percentage multigenerational household, percentage non-white, percentage non-American parent, percentage in school, percentage unable to read and write, percentage unemployed, percentage in labour force, percentage in each industry

Table 14 Balance table showing the 1910 characteristics of the matched men in 1930

	Treatment	Comparison	Difference	t-stat	*p* value
Individual					
Non-white (%)	1.86	0.48	− 1.38	− 5.23	0.00***
Non-American born parent (%)	29.58	27.79	− 1.80	− 1.24	0.21
Household (1910)					
Family size	5.98	5.95	− 0.02	− 0.32	0.75
Families in household	1.26	1.26	0.00	0.18	0.85
Non-family household (%)	0.35	0.34	− 0.01	− 0.05	0.96
Urban household (%)	17.27	17.03	− 0.25	− 0.20	0.84
Farm household (%)	52.97	60.31	7.34	4.65	0.00***
Paying mortgage (%)	29.05	36.69	7.64	4.96	0.00***
Mother (1910)					
Age	31.70	31.34	− 0.35	− 1.60	0.11
Non-American born parent (%)	44.29	46.82	2.54	1.58	0.11
Father (1910)					
Age	38.33	37.10	− 1.24	− 4.81	0.00***
Non-American born parent (%)	45.08	49.09	4.00	2.49	0.01*
Education score	9.88	8.35	− 1.52	− 3.74	0.00***
Earnings score	34.86	29.06	− 5.79	− 6.04	0.00***
Unemployed (%)	5.31	5.01	− 0.31	− 0.44	0.66

Notes: (i) stars indicate significance according to $*p < 0.05$, $**p < 0.01$, $***p < 0.001$; (ii) the population variables are obtained from the Integrated Public Use Microdata Series USA 1910 and 1930 full-count individual censuses

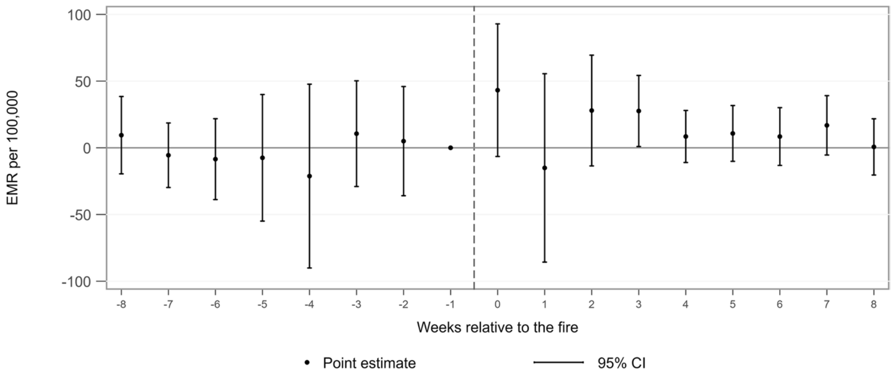


Fig. 10 Difference-in-Differences results regressing the excess mortality rate of children under the age of five on burned area (weeks 0–8). *Notes:* (i) this plot shows the coefficient of the excess mortality rate per 100,000 children regressed on the burned area indicator variable in Eq. (2); (ii) EMR denotes the excess mortality rate and CI indicates the 95% confidence interval; (iii) the mortality data are obtained from the genealogy company Ancestry.com and the population data are retrieved from the 1910 US census

Table 15 Difference-in-Differences regression results of smoke exposure on the excess mortality rate of children under the age of five (weeks 0–8)

	Excess mortality rate		
	(1)	(2)	(3)
Smoke event	51.1	64.4*	35.0*
	(27.8)	(29.3)	(15.8)
1 week after event	48.3	45.7	55.9
	(30.1)	(33.7)	(33.8)
2 weeks after event	22.4	28.8	9.8
	(23.4)	(26.6)	(14.3)
3 weeks after event	2.0	9.7	− 9.0
	(9.6)	(13.3)	(11.4)
4 weeks after event	− 2.1	8.3	2.6
	(9.8)	(12.8)	(7.9)
5 weeks after event	11.6	4.4	− 2.9
	(10.5)	(15.3)	(9.7)
6 weeks after event	2.6	19.7	14.0
	(7.2)	(11.0)	(10.1)
7 weeks after event	18.0	25.8	14.4
	(11.5)	(14.9)	(9.2)
8 weeks after event	0.6	9.1	8.6
	(9.5)	(12.2)	(8.9)
County fixed effects	✓	✓	✓
Week fixed effects		✓	✓
Burned area			✓
R^2	0.09	0.11	0.11
N	1190	1190	1190

Notes: (i) stars indicate significance according to $*p < 0.05$, $**p < 0.01$, $***p < 0.001$; (ii) the table shows the coefficients of the Difference-in-Differences estimation as stated in Eq. (2); (iii) the sample includes 70 counties; (iv) Burned area indicates whether a county comprised burned area; (v) standard errors are clustered at the county level; (vi) the population data are compiled from the 1910 US full-count census provided by the Integrated Public Use Microdata Series (IPUMS) USA and the mortality data are retrieved from the genealogy company Ancestry.com

Table 16 Difference-in-Differences regression results of smoke exposure on the excess mortality rate of children under the age of five (weeks 0 to 4) excluding the fire-affected counties

	Excess mortality rate	
	(1)	(2)
Smoke event	6.0	19.8*
	(3.1)	(9.6)
1 week after event	66.6	64.0
	(42.1)	(45.2)
2 weeks after event	− 5.3	1.5
	(2.7)	(13.2)
3 weeks after event	− 9.2	0.1
	(4.8)	(10.6)
4 weeks after event	− 2.9	8.2
	(1.5)	(8.5)
County fixed effects	✓	✓
Week fixed effects		✓
R^2	0.09	0.11
N	986	986

Notes: (i) stars indicate significance according to $*p < 0.05$, $**p < 0.01$, $***p < 0.001$; (ii) the table shows the coefficients of the Difference-in-Differences estimation as stated in Eq. (2); (iii) the sample includes 58 counties; (iv) the population data are compiled from the 1910 US full-count census provided by the Integrated Public Use Microdata Series (IPUMS) USA and the mortality data are retrieved from the genealogy company Ancestry.com

Table 17 Regression of individual socioeconomic status outcomes in 1930 on smoke exposure in early childhood excluding fire-affected counties

	Composite		Single dimension		
	(1)	(2)	(3)	(4)	(5)
	SEI	NPBOSS	OCC.INC	OCC.ER	OCC.ED
Smoke	− 8.0***	− 8.4**	− 3.5**	− 9.9**	− 1.6
	(2.2)	(3.1)	(1.2)	(3.5)	(2.0)
Non-white	− 9.9***	− 19.7***	− 6.8***	− 19.6***	− 5.7***
	(2.1)	(3.0)	(1.2)	(4.2)	(0.8)
Non-American born parent	− 0.5	− 0.8	− 0.2	− 0.6	− 0.4
	(0.7)	(1.0)	(0.4)	(1.2)	(0.7)
Household (1910)					
Family size	− 0.8***	− 0.8***	− 0.2***	− 0.5**	− 0.6***
	(0.1)	(0.2)	(0.1)	(0.2)	(0.1)
Families in household	0.1	− 0.1	− 0.1	− 0.3	− 0.1
	(0.3)	(0.4)	(0.2)	(0.6)	(0.3)
Non-family household	3.1	4.3	1.3	5.6	4.2
	(4.2)	(6.2)	(2.4)	(7.1)	(4.2)
Urban household	4.5***	4.1***	1.1*	2.4	1.9**
	(0.8)	(1.1)	(0.4)	(1.5)	(0.6)
Farm household	− 2.3*	− 4.9***	− 1.8***	− 5.4***	− 1.5*
	(0.9)	(1.3)	(0.5)	(1.4)	(0.7)
Paying mortgage	0.6	0.7	0.1	0.5	0.7
	(0.5)	(0.6)	(0.2)	(0.6)	(0.4)
Parents (1910)					
Mother: Age	0.3***	0.4***	0.1***	0.3***	0.2***
	(0.0)	(0.1)	(0.0)	(0.1)	(0.1)
Mother: Non-American born parent	− 0.4	− 0.8	− 0.5	− 1.4	− 0.2
	(0.5)	(0.7)	(0.3)	(0.8)	(0.4)
Father: Age	− 0.1	− 0.1	− 0.0	− 0.1	− 0.0
	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)
Father: Non-American born parent	0.8	− 0.2	− 0.3	− 1.1	0.2
	(0.6)	(0.8)	(0.3)	(0.8)	(0.5)
Father: Education score	0.1***	0.1	0.0	− 0.0	0.1**
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Earnings score	0.1**	0.1***	0.0***	0.1***	0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Unemployed	− 0.9	− 0.9	− 0.3	− 0.7	− 1.3*
	(0.9)	(1.2)	(0.4)	(1.1)	(0.6)
Controls					
Industry father	✓	✓	✓	✓	✓
County characteristics	✓	✓	✓	✓	✓
State indicator	✓	✓	✓	✓	✓
R^2	0.12	0.15	0.13	0.14	0.06

Table 17 (continued)

	Composite		Single dimension		
	(1)	(2)	(3)	(4)	(5)
	SEI	NPBOSS	OCC.INC	OCC.ER	OCC.ED
N	6505	6455	6505	6455	6438

Notes: (i) stars indicate significance according to $*p < 0.05$, $**p < 0.01$, $***p < 0.001$; (ii) this table shows the results of the Ordinary Least Squares regression shown in Eq. (3) estimating the effect of wildfire smoke exposure in early childhood on later-life socioeconomic status conditional on controls; (iii) the data are obtained from the Integrated Public Use Microdata Series full-count censuses 1910 and 1930; (iv) the composite socioeconomic status measure SEI stands for the Duncan Socioeconomic Index, and NPBOSS is the Nam–Powers–Boyd Occupational Status Score; (v) the single dimension socioeconomic status measures are occupational income (OCC.INC), earnings (OCC.ER), and education (OCC.ED)

Table 18 Regression of individual socioeconomic status outcomes in 1940 on smoke exposure in early childhood excluding fire-affected counties

	Composite		Single dimension		
	(1)	(2)	(3)	(4)	(5)
	SEI	NPBOSS	OCC.INC	OCC.ER	OCC.ED
Smoke	− 9.6*	− 9.5*	− 5.8***	− 12.7**	− 7.4
	(4.0)	(4.5)	(1.5)	(4.4)	(4.1)
Non-white	− 11.7**	− 12.5*	− 3.6	− 9.8	− 6.2
	(4.3)	(5.3)	(1.9)	(6.0)	(3.4)
Non-American born parent	− 1.2	− 2.4	− 0.8	− 2.5	− 0.5
	(1.1)	(1.3)	(0.5)	(1.4)	(1.0)
Household (1910)					
Family size	− 1.1***	− 1.0***	− 0.3**	− 0.7**	− 0.9***
	(0.2)	(0.2)	(0.1)	(0.2)	(0.2)
Families in household	0.0	0.1	0.0	0.0	− 0.3
	(0.5)	(0.6)	(0.2)	(0.6)	(0.5)
Non-family household	2.0	4.3	0.4	− 0.1	2.5
	(5.3)	(5.7)	(1.9)	(6.3)	(5.5)
Urban household	4.7**	3.7*	1.4*	1.5	3.6**
	(1.5)	(1.4)	(0.5)	(1.4)	(1.3)
Farm household	− 0.2	− 1.9	− 0.5	− 2.2	− 0.1
	(1.5)	(1.5)	(0.6)	(1.4)	(1.2)
Paying mortgage	− 1.2	− 1.0	− 0.6	− 1.3	− 1.1
	(0.6)	(0.9)	(0.3)	(1.0)	(0.7)
Parents (1910)					
Mother: Age	0.3**	0.2*	0.1**	0.2	0.2
	(0.1)	(0.1)	(0.0)	(0.1)	(0.1)
Mother: Non-American born parent	− 0.3	− 1.3	− 0.2	− 1.3	− 0.0
	(0.7)	(0.9)	(0.3)	(0.9)	(0.7)
Father: Age	− 0.0	− 0.0	− 0.0	− 0.0	0.0
	(0.1)	(0.1)	(0.0)	(0.1)	(0.1)
Father: Non-American born parent	0.1	− 0.8	− 0.1	− 0.9	− 0.6
	(0.8)	(1.0)	(0.4)	(1.1)	(0.8)
Father: Education score	0.3***	0.2***	0.1***	0.2***	0.3***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Earnings score	0.1*	0.1***	0.0**	0.1**	0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Unemployed	2.1	1.0	0.1	− 0.3	2.5
	(1.7)	(2.0)	(0.7)	(2.1)	(1.7)
Controls					
Industry father	✓	✓	✓	✓	✓
County characteristics	✓	✓	✓	✓	✓
State indicator	✓	✓	✓	✓	✓
R^2	0.14	0.14	0.12	0.11	0.10

Table 18 (continued)

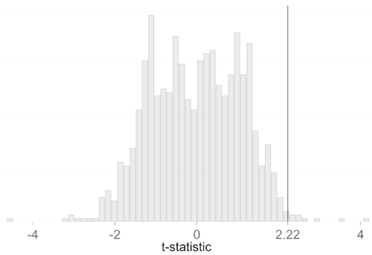
	Composite		Single dimension		
	(1)	(2)	(3)	(4)	(5)
	SEI	NPBOSS	OCC.INC	OCC.ER	OCC.ED
N	4,654	4,718	4,731	4,718	4,706

Notes: (i) stars indicate significance according to $*p < 0.05$, $**p < 0.01$, $***p < 0.001$; (ii) this table shows the results of the Ordinary Least Squares regression shown in Eq. (3) estimating the effect of wildfire smoke exposure in early childhood on later-life socioeconomic status conditional on controls; (iii) the data are obtained from the Integrated Public Use Microdata Series full-count censuses 1910 and 1940; (iv) the composite socioeconomic status measure SEI stands for the Duncan Socioeconomic Index, and NPBOSS is the Nam–Powers–Boyd Occupational Status Score; (v) the single dimension socioeconomic status measures are occupational income (OCC.INC), earnings (OCC.ER), and education (OCC.ED)

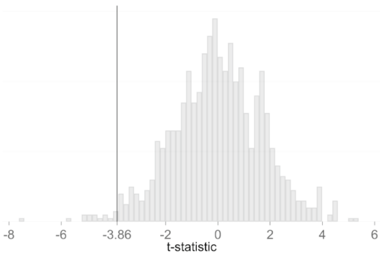
B. Robustness checks

See Figs. 11, 12, 13, 14, 15 and 16 and Tables 19 and 20.

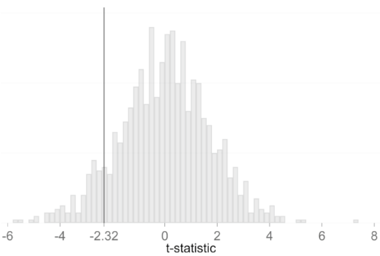
See Fig. 17 and Tables 21, 22, 23 and 24.



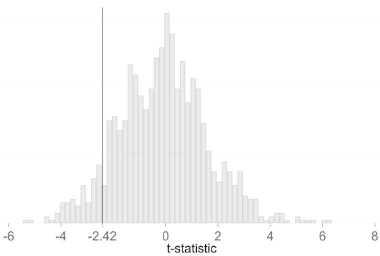
(a) Excess mortality rate in the fire week



(b) Duncan Socioeconomic Index 1930

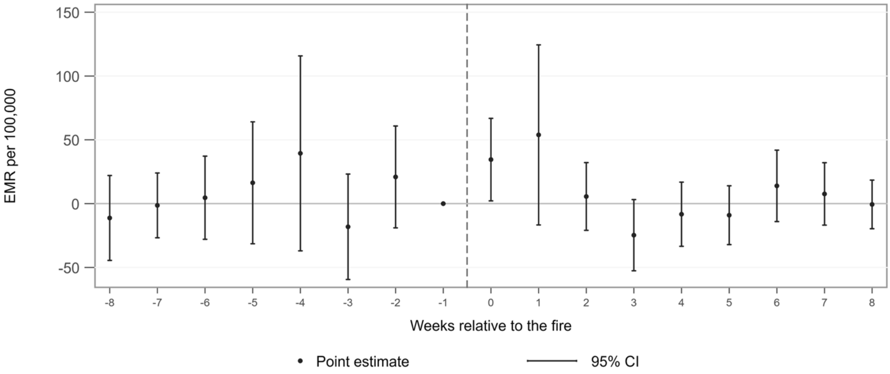


(c) Nam-Powers-Boyd Occupational Status Score 1930

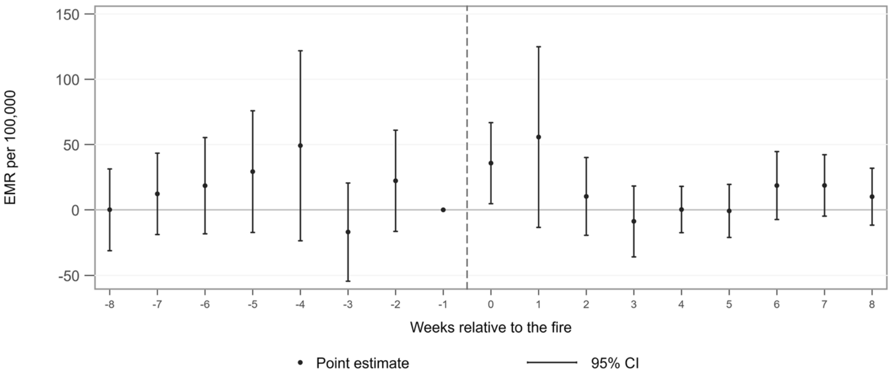


(d) Occupational Income 1940

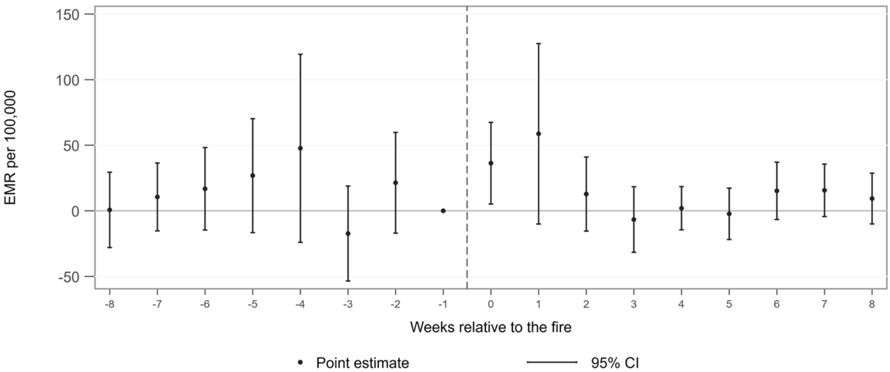
Fig. 11 Permutation test of the significant effects of smoke exposure on excess mortality and later-life socioeconomic status indices with 1000 iterations. *Notes:* (i) **a** shows the distribution of the t-statistic for 1000 iterations of Eq. (2) (short-run estimation) randomly assigning treatment (smoke exposure) to 14 out of the 70 counties of which 10 are also containing burned area (ii) **b**, **c**, and **d** show the distribution of the t-statistic for 1000 iterations of Eq. (3) (long-run estimation) randomly assigning treatment (smoke exposure) to 14 out of the 70 counties of which 10 are also containing burned area; (iii) the vertical black lines indicates the t-statistic of the baseline estimates presented in Table 4 Column (3) for (**a**), of Table 5 Columns (1) and (2) for (**b**) and (**c**) in 1930, respectively, and of Table 6 Column (3) for (**d**)



(a) 2-year baseline period (1908-1909)

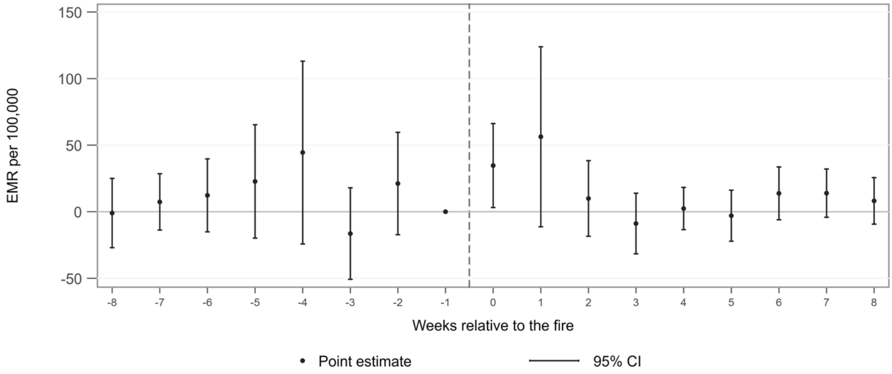


(b) 3-year baseline period (1907-1909)

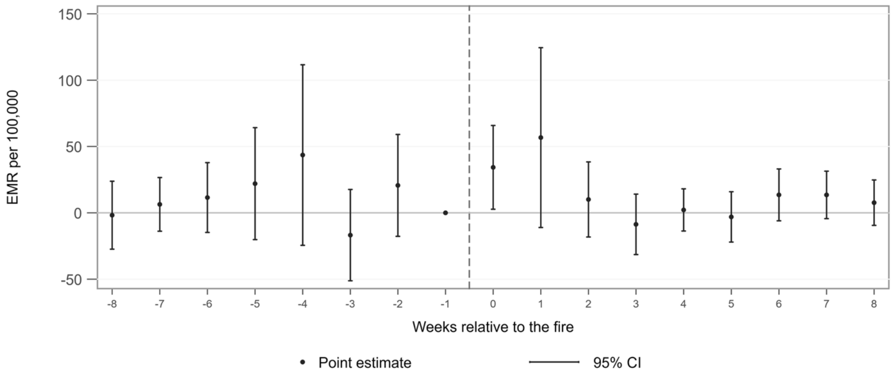


(c) 4-year baseline period (1906-1909)

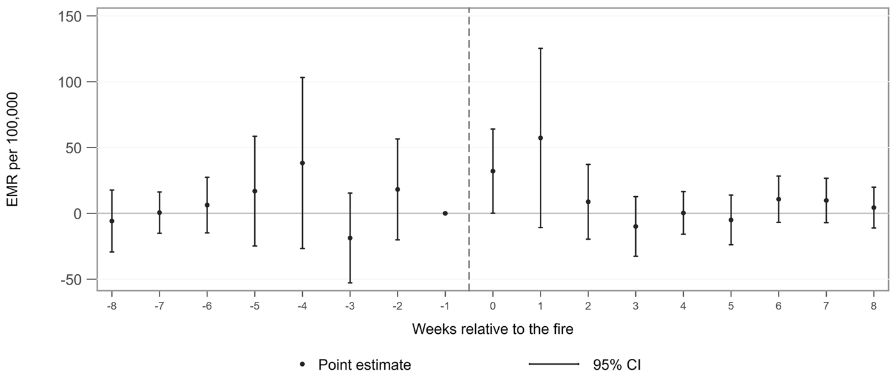
Fig. 12 Sensitivity of the excess mortality rate to the choice of the mortality baseline (reference) period. *Notes:* (i) shows the coefficient sensitivity of the excess mortality rate estimated in Eq. (2) to different reference periods; (ii) the baseline (reference) period used to compute the excess mortality rate is 2 years (**a**), 3 years (**b**), and 4 years (**c**) prior to the Great Fire



(a) Linear decrease by 5%



(b) Linear decrease by 10%



(c) Decrease by 50%

Fig. 13 Sensitivity of the excess mortality rate be reducing the weight for older years in the baseline (reference) period. *Notes:* (i) shows the coefficient sensitivity of the excess mortality rate estimated in Eq. (2) to down weighing for older years in the reference period; (ii) the weights adopted in **a** decrease linearly in 5% steps prior to 1910 (i.e. 100% for 1909, 95% for 1908,..., 80% for 1905), **b** shows a linear decrease by 10%, and in **c** the weight for each year decreases by half (i.e. 100% for 1909, 50% for 1908, etc.)

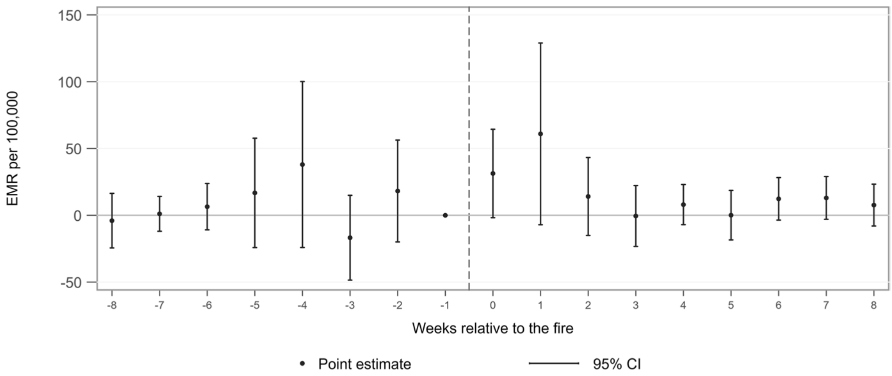
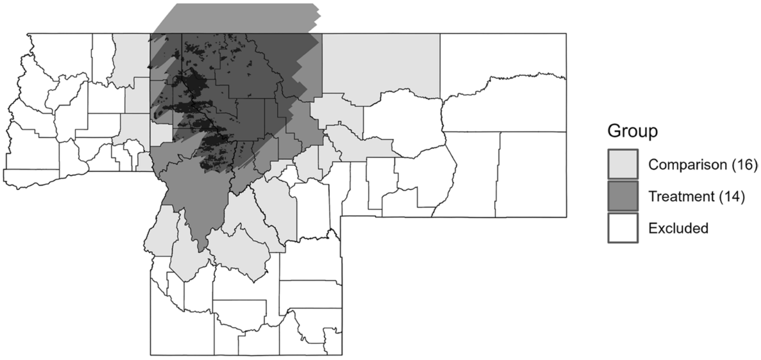
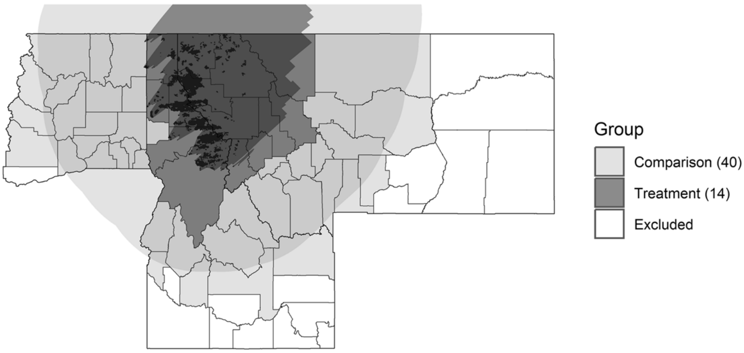


Fig. 14 Excess mortality rates estimated using a Quasi-Poisson model. *Notes:* (i) the above plot shows the coefficient of the excess mortality rate per 100,000 children using estimated Eq. (2) using a Quasi-Poisson model; (ii) the model is a Quasi-Poisson generalised linear model fitted using the R-package `glm` including a time trend variable

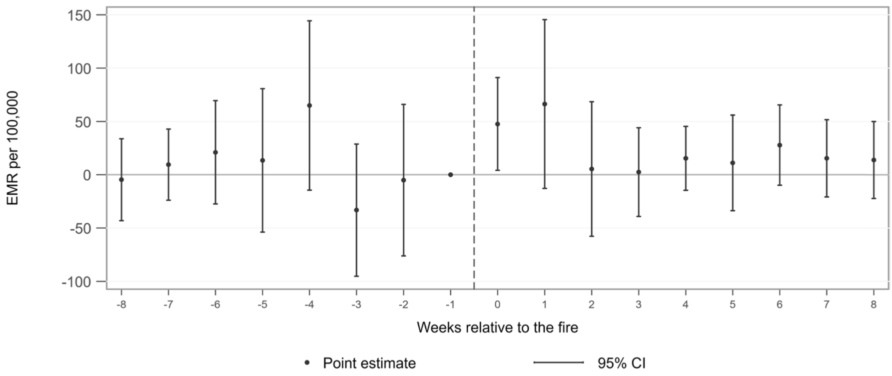


(a) Adjacent counties

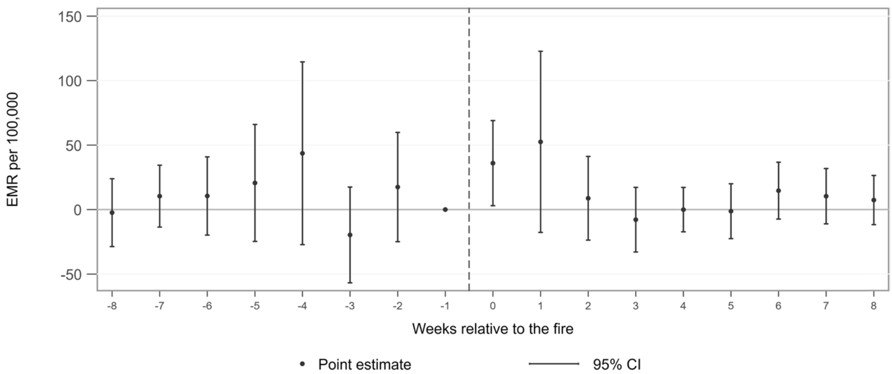


(b) Within 250 kilometre buffer

Fig. 15 Treatment and comparison group for smoke-affected counties restricting the comparison group to adjacent and surrounding counties. *Notes:* (i) a county is classified as smoke affected if any part of the county was exposed to moderate hourly peak pollution ($PM_{2.5}$ > 38 μg/m^3); (ii) the county shape file shows the historical county boundaries provided by The Big Ten Academic Alliance Geoportal; (iii) the lighter grey area shows the moderate hourly peak pollution area and the darker shaded grey area indicates the 250 km buffer around this area



(a) Adjacent counties



(b) Within 250 kilometre buffer

Fig. 16 Difference-in-Differences point estimates and 95% confidence intervals of smoke exposure on excess mortality of children under the age of five only including comparison counties surrounding smoke-affected counties. *Notes:* (i) EMR denotes the excess mortality rate and CI indicates the 95% confidence intervals; (ii) the mortality data are obtained from the genealogy company Ancestry.com and the population data are retrieved from the 1910 US census

Table 19 Difference-in-Differences regression results of smoke exposure on the excess mortality rate of children under the age of five (weeks 0 to 8) only including 16 adjacent counties

	Excess mortality rate		
	(1)	(2)	(3)
Smoke event	51.1	77.4*	47.6*
	(28.4)	(35.1)	(21.3)
1 week after event	48.3	61.0	66.3
	(30.7)	(41.1)	(38.7)
2 weeks after event	22.4	20.3	5.3
	(23.8)	(44.0)	(30.9)
3 weeks after event	2.0	22.7	2.5
	(9.8)	(24.3)	(20.4)
4 weeks after event	− 2.1	24.8	15.4
	(10.0)	(19.1)	(14.7)
5 weeks after event	11.6	22.3	11.1
	(10.7)	(29.9)	(21.9)
6 weeks after event	2.6	37.5	27.8
	(7.3)	(20.5)	(18.4)
7 weeks after event	18.0	25.7	15.4
	(11.8)	(24.6)	(17.7)
8 weeks after event	0.6	16.1	13.9
	(9.7)	(22.6)	(17.6)
County fixed effects	✓	✓	✓
Week fixed effects		✓	✓
Burned area			✓
R^2	0.11	0.16	0.17
N	510	510	510

Notes: (i) stars indicate significance according to $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; (ii) the table shows the coefficients of the Difference-in-Differences estimation as stated in Eq. (2); (iii) the sample includes 30 counties; (iv) Burned area indicates whether a county comprised burned area; (v) standard errors are clustered at the county level; (vi) the population data are compiled from the 1910 US full-count census provided by the Integrated Public Use Microdata Series (IPUMS) USA and the mortality data are retrieved from the genealogy company Ancestry.com

Table 20 Difference-in-Differences regression results of smoke exposure on the excess mortality rate of children under the age of five (weeks 0–8) only including 40 surrounding comparison courtiers within a 250 km buffer

	Excess mortality rate		
	(1)	(2)	(3)
Smoke event	51.1	65.1*	36.0*
	(27.9)	(30.0)	(16.5)
1 week after event	48.3	41.2	52.5
	(30.2)	(36.0)	(35.0)
2 weeks after event	22.4	26.9	8.7
	(23.5)	(28.4)	(16.2)
3 weeks after event	2.0	10.9	− 7.9
	(9.7)	(14.7)	(12.5)
4 weeks after event	− 2.1	4.6	− 0.1
	(9.9)	(13.7)	(8.6)
5 weeks after event	11.6	6.5	− 1.3
	(10.5)	(16.4)	(10.6)
6 weeks after event	2.6	20.5	14.7
	(7.2)	(12.3)	(11.0)
7 weeks after event	18.0	20.0	10.4
	(11.6)	(16.6)	(10.7)
8 weeks after event	0.6	7.3	7.4
	(9.5)	(13.1)	(9.5)
County fixed effects	✓	✓	✓
Week fixed effects		✓	✓
Burned area			✓
R^2	0.09	0.11	0.12
N	918	918	918

Notes: (i) stars indicate significance according to $*p < 0.05$, $**p < 0.01$, $***p < 0.001$; (ii) the table shows the coefficients of the Difference-in-Differences estimation as stated in Eq. (2); (iii) the sample includes 54 counties; (iv) Burned area indicates whether a county comprised burned area; (v) standard errors are clustered at the county level; (vi) the population data are compiled from the 1910 US full-count census provided by the Integrated Public Use Microdata Series (IPUMS) USA and the mortality data are retrieved from the genealogy company Ancestry.com

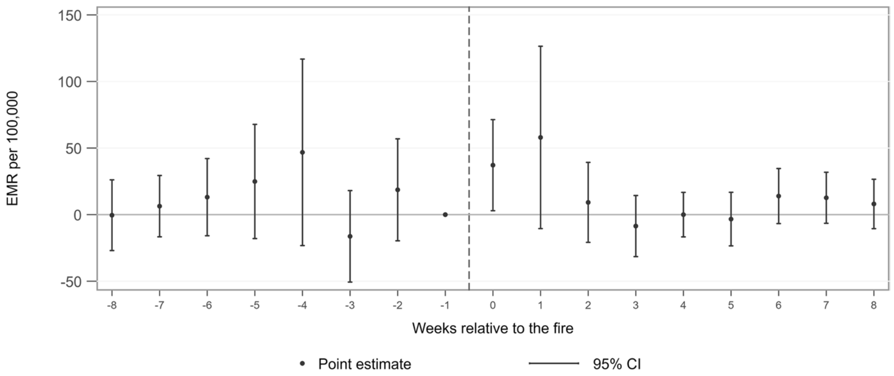


Fig. 17 Difference-in-Differences point estimates and 95% confidence intervals of smoke exposure on excess mortality of children under the age of five excluding Spokane County and Missoula County. *Notes:* (i) EMR denotes the excess mortality rate and CI indicates the 95% confidence interval; (ii) the mortality data are obtained from the genealogy company Ancestry.com and the population data are retrieved from the 1910 US census

Table 21 Difference-in-Differences regression results of smoke exposure on the excess mortality rate of children under the age of five (weeks 0 to 8) excluding Spokane County and Missoula County

	Excess mortality rate		
	(1)	(2)	(3)
Smoke event	55.4	68.3*	37.1*
	(29.6)	(31.1)	(17.1)
1 week after event	53.0	50.6	58.0
	(32.0)	(35.5)	(34.3)
2 weeks after event	20.6	26.8	9.2
	(25.1)	(28.3)	(15.0)
3 weeks after event	2.5	10.1	− 8.6
	(10.4)	(14.0)	(11.5)
4 weeks after event	− 7.5	2.5	− 0.0
	(9.0)	(12.2)	(8.4)
5 weeks after event	11.5	3.6	− 3.4
	(11.3)	(16.0)	(10.1)
6 weeks after event	1.8	19.2	13.9
	(7.7)	(11.5)	(10.4)
7 weeks after event	13.8	21.6	12.6
	(11.6)	(15.1)	(9.6)
8 weeks after event	− 0.4	7.8	8.0
	(10.2)	(12.9)	(9.3)
County fixed effects	✓	✓	✓
Week fixed effects		✓	✓
Burned area			✓
R^2	0.09	0.11	0.12
N	1156	1156	1156

Notes: (i) stars indicate significance according to $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; (ii) the table shows the coefficients of the Difference-in-Differences estimation as stated in Eq. (2); (iii) the sample includes 68 counties; (iv) Burned area indicates whether a county comprised burned area; (v) standard errors are clustered at the county level; (vi) the population data are compiled from the 1910 US full-count census provided by the Integrated Public Use Microdata Series (IPUMS) USA and the mortality data are retrieved from the genealogy company Ancestry.com

Table 22 Regression of individual socioeconomic status outcomes in 1930 on smoke exposure in early childhood excluding Spokane County and Missoula County

	Composite		Single dimension		
	(1)	(2)	(3)	(4)	(5)
	SEI	NPBOSS	OCC.INC	OCC.ER	OCC.ED
Smoke	− 2.5**	− 1.9	− 0.5	− 1.1	− 0.5
	(0.8)	(1.1)	(0.4)	(1.3)	(0.5)
Burned area	2.9*	0.4	− 0.4	− 2.6	0.1
	(1.2)	(1.5)	(0.6)	(1.5)	(0.7)
Individual					
Non-white	− 8.2***	− 17.0***	− 6.2***	− 17.8***	− 4.0**
	(2.1)	(3.1)	(1.1)	(3.6)	(1.2)
Non-American born parent	− 1.0	− 1.2	− 0.3	− 1.0	− 0.7
	(0.7)	(1.1)	(0.4)	(1.2)	(0.6)
Household (1910)					
Family size	− 0.9***	− 0.9***	− 0.3***	− 0.6**	− 0.6***
	(0.1)	(0.2)	(0.1)	(0.2)	(0.1)
Families in household	0.3	0.1	− 0.1	− 0.2	− 0.0
	(0.3)	(0.4)	(0.2)	(0.5)	(0.2)
Non-family household	2.8	3.4	1.3	5.4	3.1
	(3.9)	(6.0)	(2.4)	(7.1)	(3.9)
Urban household	4.7***	4.8***	1.4**	3.5*	2.3**
	(0.9)	(1.2)	(0.5)	(1.6)	(0.7)
Farm household	− 2.5**	− 5.1***	− 1.8***	− 5.5***	− 1.1
	(0.8)	(1.2)	(0.4)	(1.2)	(0.7)
Paying mortgage	0.7	0.8	0.1	0.6	0.7
	(0.4)	(0.6)	(0.2)	(0.6)	(0.4)
Parents (1910)					
Mother: Age	0.3***	0.4***	0.1***	0.3***	0.2***
	(0.0)	(0.1)	(0.0)	(0.1)	(0.1)
Mother: Non-American born parent	− 0.1	− 0.4	− 0.3	− 1.0	− 0.1
	(0.5)	(0.7)	(0.3)	(0.8)	(0.4)
Father: Age	− 0.1	− 0.1	− 0.0	− 0.1	− 0.0
	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)
Father: Non-American born parent	0.5	− 0.7	− 0.4	− 1.4	0.1
	(0.6)	(0.8)	(0.3)	(0.8)	(0.5)
Father: Education score	0.1**	0.1	0.0	0.0	0.1**
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Earnings score	0.1**	0.1***	0.0***	0.1***	0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Unemployed	− 1.3	− 1.6	− 0.4	− 1.2	− 1.5*
	(0.9)	(1.3)	(0.4)	(1.2)	(0.6)
Controls					
Industry father	✓	✓	✓	✓	✓

Table 22 (continued)

	Composite		Single dimension		
	(1)	(2)	(3)	(4)	(5)
	SEI	NPBOSS	OCC.INC	OCC.ER	OCC.ED
County characteristics	✓	✓	✓	✓	✓
State indicator	✓	✓	✓	✓	✓
R^2	0.10	0.13	0.12	0.12	0.05
N	7044	6988	7044	6988	6963

Notes: (i) stars indicate significance according to $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; (ii) this table shows the results of the Ordinary Least Squares regression shown in Eq. (3) estimating the effect of wildfire smoke exposure in early childhood on later-life socioeconomic status conditional on controls; (iii) the data are obtained from the Integrated Public Use Microdata Series full-count censuses 1910 and 1930; (iv) the composite socioeconomic status measure SEI stands for the Duncan Socioeconomic Index, and NPBOSS is the Nam–Powers–Boyd Occupational Status Score; (v) the single dimension socioeconomic status measures are occupational income (OCC.INC), earnings (OCC.ER), and education (OCC.ED)

Table 23 Regression of individual socioeconomic status outcomes in 1940 on smoke exposure in early childhood excluding Spokane County and Missoula County

	Composite		Single dimension		
	(1)	(2)	(3)	(4)	(5)
	SEI	NPBOSS	OCC.INC	OCC.ER	OCC.ED
Smoke	0.7	– 0.2	– 0.7	– 1.4	1.2
	(0.9)	(1.0)	(0.4)	(1.1)	(0.9)
Burned area	– 0.2	– 0.3	0.3	0.2	– 0.3
	(1.1)	(1.4)	(0.5)	(1.6)	(1.1)
Individual					
Non-white	– 10.7***	– 14.0***	– 3.5*	– 10.5**	– 3.9
	(3.1)	(3.3)	(1.4)	(3.7)	(3.0)
Non-American born parent	– 0.9	– 1.6	– 0.5	– 1.7	– 0.6
	(0.7)	(0.8)	(0.3)	(1.0)	(0.6)
Household (1910)					
Family size	– 1.1***	– 1.0***	– 0.3***	– 0.7***	– 0.8***
	(0.1)	(0.2)	(0.1)	(0.2)	(0.1)
Families in household	0.7	0.7	0.3	0.7	0.2
	(0.4)	(0.4)	(0.2)	(0.4)	(0.4)
Non-family household	3.2	4.1	0.7	2.0	– 0.1
	(4.6)	(5.3)	(2.3)	(6.7)	(3.4)
Urban household	3.1**	2.8*	1.0*	1.4	2.5*
	(1.1)	(1.2)	(0.5)	(1.2)	(1.1)
Farm household	– 1.7	– 3.2**	– 1.2**	– 3.5**	– 0.6
	(0.9)	(1.1)	(0.4)	(1.1)	(0.8)
Paying mortgage	– 0.4	0.2	– 0.1	0.1	– 0.7
	(0.5)	(0.6)	(0.2)	(0.7)	(0.6)
Parents (1910)					
Mother: Age	0.3***	0.4***	0.1***	0.2**	0.2**
	(0.1)	(0.1)	(0.0)	(0.1)	(0.1)
Mother: Non-American born parent	0.4	– 0.3	0.2	– 0.1	0.2
	(0.5)	(0.7)	(0.2)	(0.7)	(0.5)
Father: Age	– 0.1	-0.1	– 0.0	– 0.1	– 0.0
	(0.1)	(0.1)	(0.0)	(0.1)	(0.0)
Father: Non-American born parent	0.2	– 0.9	– 0.3	– 1.3	– 0.2
	(0.6)	(0.8)	(0.3)	(0.9)	(0.6)
Father: Education score	0.2***	0.2***	0.1***	0.1***	0.3***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Earnings score	0.1***	0.1***	0.0***	0.1***	0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Unemployed	1.2	1.7	0.4	1.4	0.8
	(1.2)	(1.3)	(0.5)	(1.5)	(1.1)
Controls					
Industry father	✓	✓	✓	✓	✓

Table 23 (continued)

	Composite		Single dimension		
	(1)	(2)	(3)	(4)	(5)
	SEI	NPBOSS	OCC.INC	OCC.ER	OCC.ED
County characteristics	✓	✓	✓	✓	✓
State indicator	✓	✓	✓	✓	✓
R^2	0.11	0.11	0.09	0.08	0.08
N	8103	8213	8233	8213	8192

Notes: (i) stars indicate significance according to $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; (ii) this table shows the results of the Ordinary Least Squares regression shown in Eq. (3) estimating the effect of wildfire smoke exposure in early childhood on later-life socioeconomic status conditional on controls; (iii) the data are obtained from the Integrated Public Use Microdata Series full-count censuses 1910 and 1940; (iv) the composite socioeconomic status measure SEI stands for the Duncan Socioeconomic Index, and NPBOSS is the Nam–Powers–Boyd Occupational Status Score; (v) the single dimension socioeconomic status measures are occupational income (OCC.INC), earnings (OCC.ER), and education (OCC.ED)

Table 24 Regression of individual socioeconomic status outcomes in 1940 on smoke exposure in early childhood retaining the 1930 cohort

	Composite		Single dimension		
	(1)	(2)	(3)	(4)	(5)
	SEI	NPBOSS	OCC.INC	OCC.ER	OCC.ED
Smoke	– 0.1	– 1.2	– 1.3*	– 3.6*	2.4
	(1.3)	(1.6)	(0.5)	(1.6)	(1.3)
Burned area	1.6	2.7	1.6*	4.5*	– 0.7
	(1.3)	(2.0)	(0.6)	(2.1)	(1.5)
Individual					
Non-white	– 12.9***	– 14.3***	– 4.2**	– 10.9**	– 7.1**
	(2.9)	(3.6)	(1.3)	(4.1)	(2.2)
Non-American born parent	– 0.7	– 1.7	– 0.5	– 1.8	– 0.2
	(1.0)	(1.2)	(0.4)	(1.3)	(0.8)
Household (1910)					
Family size	– 1.2***	– 1.1***	– 0.3***	– 0.8***	– 0.9***
	(0.2)	(0.2)	(0.1)	(0.2)	(0.2)
Families in household	0.2	0.3	0.1	0.3	– 0.0
	(0.4)	(0.5)	(0.2)	(0.6)	(0.4)
Non-family household	2.8	3.9	1.7	3.9	0.3
	(5.0)	(6.3)	(2.5)	(8.0)	(4.4)
Urban household	4.9***	3.9**	1.5**	1.5	4.0**
	(1.3)	(1.2)	(0.5)	(1.2)	(1.2)
Farm household	– 0.3	– 2.4	– 0.7	– 2.8*	– 0.1
	(1.3)	(1.4)	(0.5)	(1.4)	(1.1)
Paying mortgage	– 0.6	– 0.1	– 0.2	– 0.2	– 0.8
	(0.6)	(0.8)	(0.3)	(0.9)	(0.6)
Parents (1910)					
Mother: Age	0.3***	0.3**	0.1***	0.2*	0.2*
	(0.1)	(0.1)	(0.0)	(0.1)	(0.1)
Mother: Non-American born parent	– 0.0	– 0.8	0.0	– 0.9	– 0.1
	(0.7)	(0.8)	(0.3)	(0.9)	(0.7)
Father: Age	– 0.0	– 0.0	– 0.0	– 0.0	0.0
	(0.1)	(0.1)	(0.0)	(0.1)	(0.1)
Father: Non-American born parent	0.6	– 0.4	0.1	– 0.5	– 0.1
	(0.7)	(0.8)	(0.3)	(0.9)	(0.7)
Father: Education score	0.3***	0.2***	0.1***	0.1**	0.3***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Earnings score	0.1**	0.1***	0.0***	0.1***	0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Unemployed	1.9	1.1	0.3	0.3	2.5
	(1.6)	(1.8)	(0.7)	(1.9)	(1.6)
Controls					
Industry father	✓	✓	✓	✓	✓

Table 24 (continued)

	Composite		Single dimension		
	(1)	(2)	(3)	(4)	(5)
	SEI	NPBOSS	OCC.INC	OCC.ER	OCC.ED
County characteristics	✓	✓	✓	✓	✓
State indicator	✓	✓	✓	✓	✓
R^2	0.14	0.13	0.12	0.11	0.10
N	5550	5631	5645	5631	5617

Notes: (i) stars indicate significance according to $*p < 0.05$, $**p < 0.01$, $***p < 0.001$; (ii) this table shows the results of the Ordinary Least Squares regression shown in Eq. (3) estimating the effect of wildfire smoke exposure in early childhood on later-life socioeconomic status conditional on controls; (iii) the data are obtained from the Integrated Public Use Microdata Series full-count censuses 1910 and 1940; (iv) the composite socioeconomic status measure SEI stands for the Duncan Socioeconomic Index, and NPBOSS is the Nam–Powers–Boyd Occupational Status Score; (v) the single dimension socioeconomic status measures are occupational income (OCC.INC), earnings (OCC.ER), and education (OCC.ED)

References

Abadie A, Athey S, Imbens GW, Wooldridge JM (2023) When should you adjust standard errors for clustering? Q J Econ 138(1):1–35

Abramitzky R, Boustan L, Eriksson K et al (2021) Automated linking of historical data. J Econ Lit 59(3):865–918

Abramitzky R, Boustan L, Eriksson K et al (2022a) Census Linking Project: 1910–1930 Crosswalk

Abramitzky R, Boustan L, Eriksson K et al (2022b) Census Linking Project: 1910–1940 Crosswalk

Abramitzky R, Boustan LP, Eriksson K (2012) Europe's tired, poor, huddled masses: self-selection and economic outcomes in the age of mass migration. Am Econ Rev 102(5):1832–1856

Acosta RJ, Irizarry RA (2022) A flexible statistical framework for estimating excess mortality. Epidemiology 33(3):346–353

Bailey RE, Hatton TJ, Inwood K (2018) Atmospheric pollution, health, and height in late nineteenth century Britain. J Econ Hist 78(4):1210–1247

Barreca A, Clay K, Tarr J (2014) Coal, smoke, and death: bituminous coal and American home heating. Technical report, National Bureau of Economic Research

Beach B, Hanlon WW (2018) Coal smoke and mortality in an early industrial economy. Econ J 128(615):2652–2675

Benner RC, O'Connor JJ Jr (1913) The smoke nuisance; a question of conservation. Ind Eng Chem 5(7):587–593

Bennett WD, Zeman KL, Jarabek AM (2007) Nasal contribution to breathing and fine particle deposition in children versus adults. J Toxicol Environ Health Part A Curr Issues 71(3):227–237

Brimblecombe P (2011) The big smoke: a history of air pollution in London since medieval times. Routledge, London
Brockmeyer S, D'Angiulli A (2016) How air pollution alters brain development: the role of neuroinflammation. Transl Neurosci 7(1):24–30
Burke M et al (2022) Exposures and behavioural responses to wildfire smoke. Nat Hum Behav 6(10):1351–1361
Chen H, Samet JM, Bromberg PA, Tong H (2021) Cardiovascular health impacts of wildfire smoke exposure. Part Fibre Toxicol 18(1):1–22
Clay K, Lewis J, Severnini E (2016) Canary in a coal mine: infant mortality, property values, and tradeoffs associated with mid-20th century air pollution. Technical report, National Bureau of Economic Research
Collins WJ, Wanamaker MH (2014) Selection and economic gains in the great migration of African Americans: new evidence from linked census data. Am Econ J Appl Econ 6(1):220–252
Colmer JM, Qin S, Voorheis JL, Walker R (2024) The changing nature of pollution, income, and environmental inequality in the United States. NBER Working Paper 32060
Degroot D, Anchukaitis KJ, Tierney JE et al (2022) The history of climate and society: a review of the influence of climate change on the human past. Environ Res Lett 17(10):103001
Department of Commerce and Labor (1912) Mortality statistics: 1910. U.S. Census Bureau 109:1–191
Deryugina T, Reif J (2023) The long-run effect of air pollution on survival. NBER Working Paper 31858
Diaz HF, Swetnam TW (2013) The wildfires of 1910: climatology of an extreme early twentieth-century event and comparison with more recent extremes. Bull Am Meteor Soc 94(9):1361–1370
Doubleday A et al (2020) Mortality associated with wildfire smoke exposure in Washington state, 2006–2017: a case-crossover study. Environ Health Glob Access Sci Source 19(1):1–10
Dowdy AJ, Pepler A (2018) Pyroconvection risk in Australia: climatological changes in atmospheric stability and surface fire weather conditions. Geophys Res Lett 45(4):2005–2013
Duncan OD (1961) A socioeconomic index for all occupations. In: Reiss AJ (ed) Occupations and social status. Free Press, New York, pp 109–138
Egan T (2009) The Big Burn: Teddy Roosevelt & the fire that saved America. Houghton Mifflin Harcourt, New York, p 336
England P (1979) Women and occupational prestige: a case of vacuous sex equality. Signs J Women Culture Soc 5(2):252–265
Fisher RA (1937) The design of experiments, 2nd edn. Oliver & Boyd, Edinburgh
Frankopan P (2023) The earth transformed: an untold history. Bloomsbury Publishing, London
Frye D, Kagy G (2023) Economic consequences of childhood exposure to urban environmental toxins. WP, 53706
Gao Y et al (2023) Long-term impacts of non-occupational wildfire exposure on human health: a systematic review. Environ Pollut 320:121041
Goss WFM (1916) Smoke as a source of atmospheric pollution. JB Lippincott Company, Philadelphia
Grant E, Runkle JD (2022) Long-term health effects of wildfire exposure: a scoping review. J Clim Change Health 6:100110
Hänninen OO et al (2009) Population exposure to fine particles and estimated excess mortality in Finland from an East European wildfire episode. J Eposure Sci Environ Epidemiol 19(4):414–422
Harper K (2021) Plagues upon the Earth: disease and the course of human history. Princeton University Press, Princeton
Hauser RM, Warren JR (1997) Socioeconomic indexes for occupations: a review, update, and critique. Sociol Methodol 27(1):177–298
Heblich S, Trew A, Zylberberg Y (2021) East-side story: historical pollution and persistent neighborhood sorting. J Polit Econ 129(5):1508–1552
Hornbeck R, Keniston D (2017) Creative destruction: barriers to urban growth and the Great Boston Fire of 1872. Am Econo Rev 107(6):1365–1398
Johnston F, Hanigan I, Henderson S et al (2011) Extreme air pollution events from bushfires and dust storms and their association with mortality in Sydney, Australia 1994–2007. Environ Res 111(6):811–816
Karlinsky A, Kobak D (2021) Tracking excess mortality across countries during the Covid-19 pandemic with the world mortality dataset. eLife 10:1–21
Kelley EW, Baird JC, Bradeen O et al (1944) Early days in the Forest Service, vol 1
Kochi I, Champ PA, Loomis JB, Donovan GH (2012) Valuing mortality impacts of smoke exposure from major Southern California wildfires. J For Econ 18(1):61–75

Kollanus V et al (2017) Mortality due to vegetation fire—originated $PM_{2.5}$ exposure in Europe—Assessment for the years 2005 and 2008. Environ Health Perspect 125(1):30–37

Krainz TA (2012) Fleeing the big burn: refugees, informal assistance, and welfare practices in the progressive Era. J Policy Hist 24(3):405–431

Larkin NK et al (2009) The BlueSky smoke modeling framework. Int J Wildland Fire 18(8):906–920

Levitt M, Zonta F, Ioannidis JP (2023) Excess death estimates from multiverse analysis in 2009–2021. Eur J Epidemiol 38(11):1129–1139

Liu JC, Pereira G, Uhl SA et al (2015) A systematic review of the physical health impacts from non-occupational exposure to wildfire smoke. Environ Res 136:120–132

Ljungqvist FC, Seim A, Huhtamaa H (2021) Climate and society in European history. Wiley Interdiscip Rev Clim Change 12(2):e691

Long J, Ferrie J (2013) Intergenerational occupational mobility in Great Britain and the United States since 1850. Am Econ Rev 103(4):1109–1137

Mendes-Victor LA, Oliveira CS, Azevedo J et al (2009) The 1755 Lisbon earthquake and the genesis of the risk management concept. The 1755 Lisbon Earthquake: Revisited, pp 147–165

Miller D, Cohen S (2021) The Big Burn: the Northwest's great forest fire of 1910, 8th edn. Mountain Press Publishing Company, Missoula

Morita T et al (2017) Excess mortality due to indirect health effects of the 2011 triple disaster in Fukushima, Japan: a retrospective observational study. J Epidemiol Community Health 71(10):974–980

Msemburi W et al (2023) The WHO estimates of excess mortality associated with the COVID-19 pandemic. Nature 613(7942):130–137

Mullen L (2021) gender: predict gender from names using historical data. R package version 0.6.0

Nam CB, Boyd M (2004) Occupational status in 2000; over a century of census-based measurement. Popul Res Policy Rev 23(5):327–358

Noghanibehambari H, Fletcher J (2022) Dust to feed, dust to grey: the effect of in-utero exposure to the dust bowl on old-age longevity. NBER Working Paper 30531

Noufaily A et al (2012) An improved algorithm for outbreak detection in multiple surveillance systems. Stat Med 32(7):1206–1222

Pullabhotla HK, Souza M (2022) Air pollution from agricultural fires increases hypertension risk. J Environ Econ Manag 115:102723

Rangel MA, Vogl TS (2019) Agricultural fires and health at birth. Rev Econ Stat 101(4):616–630

Reynolds R (1903) Fire investigations in Marinette County, Wisconsin. Technical report

Rocha R, Sant'Anna AA (2022) Winds of fire and smoke: air pollution and health in the Brazilian Amazon. World Dev 151:105722

Rosales-Rueda M, Triyana M (2019) The persistent effects of early-life exposure to air pollution: evidence from the Indonesian forest fires. J Hum Resour 54(4):1037–1080

Ruggles S, Nelson MA, Sobek M et al (2024) IPUMS ancestry full count data: version 4.0. [1900–1940], Minneapolis. https://doi.org/10.18128/D014.V4.0

Santos-Burgoa C et al (2018) Differential and persistent risk of excess mortality from Hurricane Maria in Puerto Rico: a time-series analysis. Lancet Planet Health 2(11):e478–e488

Sapkota A et al (2005) Impact of the 2002 Canadian forest fires on particulate matter air quality in Baltimore City. Environ Sci Technol 39(1):24–32

Siodla J (2015) Razing San Francisco: the 1906 disaster as a natural experiment in urban redevelopment. J Urban Econ 89:48–61

Stradling D, Thorsheim P (1999) The smoke of great cities: British and American efforts to control air pollution, 1860–1914. Environ Hist 4(1):6–31

Suades-González E, Gascon M, Guxens M, Sunyer J (2015) Air pollution and neuropsychological development: a review of the latest evidence. Endocrinology 156(10):3473–3482

Talbott EO et al (2015) Fine particulate matter and the risk of autism spectrum disorder. Environ Res 140:414–420

Tan-Soo JS, Pattanayak SK (2019) Seeking natural capital projects: forest fires, haze, and early-life exposure in Indonesia. Proc Natl Acad Sci USA 116(12):5239–5245

van Wagtendonk JW (2007) The history and evolution of wildland fire use. Fire Ecol 3(2):3–17

Weigle W (1934) The great Idaho fire of 1910. Timberman

Authors and Affiliations

Sarah Meier[1] · **Eric Strobl**[2,3] · **Robert J. R. Elliott**[3]

Sarah Meier
s.meier@exeter.ac.uk

Eric Strobl
eric.strobl@unibe.ch

Robert J. R. Elliott
r.j.elliott@bham.ac.uk

1 Land, Environment, Economics and Policy Institute (LEEP), Department of Economics, University of Exeter, Exeter, UK

2 Department of Economics, University of Bern, Bern, Switzerland

3 Business School, Department of Economics, University of Birmingham, Birmingham, UK