ENERGYECOLAB

Air Pollution from Agricultural Fires Increases Hypertension Risk

Hemant K. Pullabhotla[‡], Mateus Souza[§]

Updated: July 2022

EEL Discussion Paper 111

In many parts of the developing world, farmers widely use deliberate fires to burn vegetation and clear land to plant crops. These agricultural fires, however, are known to be associated with health costs due to increased air pollution. We contribute to underpinning the associated health cost estimates by studying the effects of these fires on hypertension risk. Despite being one of the leading causes of mortality globally, there is little direct evidence on how hypertension risk changes with exposure to pollution from agricultural fires. To overcome common data and empirical challenges in this setting, we match blood pressure readings from nearly 784,000 individuals across India with satellite data on 1.2 million agricultural fires, wind direction realizations, and local ambient air pollution. We find that the incidence of hypertension increases by 1.8% for each standard deviation increase in the number of upwind fires observed one day before the blood pressure readings. We find that the impact is stronger among older males, smokers, individuals that were already on blood pressure medication, and individuals belonging to socially marginalized groups. Our estimates suggest that agricultural fires in India lead to hypertension-related additional mortality, associated with USD 9 billion annually in costs.

Keywords: air pollution, agricultural fires, cardiovascular health, hypertension

[‡] Center on Food Security and The Environment (FSE), Stanford University; Department of Economics, Deakin University

[§] EnergyEcoLab, Department of Economics, Universidad Carlos III de Madrid

http://energyecolab.uc3m.es/

uc3m | Universidad Carlos III de Madrid
DEPARTAMENTO DE ECONOMÍA

ENERGYECOLAB

Non-Technical Summary

Air Pollution from Agricultural Fires Increases Hypertension Risk

The global costs of ambient air pollution amount to $6.43 trillion annually, according to recent estimates. An extensive body of work has found adverse effects of air pollution, such as the increased risk of mortality amongst children and adults, and reduced birth weights. Epidemiological evidence also suggests that health damages vary by air pollution source and composition. However, most empirical studies focus on emissions from urban or industrial sources, with limited attention given to the effects of polluting sources in rural areas. It is therefore unclear whether estimates from early studies are applicable globally, across diverse contexts and different sources of pollution. Accurate quantification of the costs of air pollution, and thus the quantification of potential benefits from its abatement, necessarily needs to account for many sources of heterogeneity.

A key challenge limiting the coverage of past studies has been the paucity of air pollution and health data in developing countries, particularly in rural areas. Coverage of air pollution monitors is sparse in many developing countries and limited to large cities or close to industrial areas. As a result, evidence on the impact of significant non-urban sources of pollution such as biomass burning, common across rural areas, is minimal. A second empirical challenge, not unique to developing countries, is finding adequate methods to deal with confounding factors. Exposure to air pollution is likely correlated with various unobserved socio-economic dimensions, behavioral patterns, or other environmental factors that also affect health outcomes. Not accounting for such unobserved factors would bias the estimate of the impact of air pollution on health.

In this study, we overcome these data and empirical challenges to provide robust evidence quantifying the health impacts of short-term air pollution exposure from rural emission in the context of a developing country. We use novel high-resolution satellite data on agricultural fires to generate plausibly exogenous variation in exposure to pollution across India. Like many other parts of the developing world, farmers in India widely use deliberate fires to burn vegetation and clear land to plant crops. During peak fire seasons in India, these agricultural fires can contribute to more than half of the particulate pollution load even within urban areas.

We focus on identifying the effects of these fires on hypertension risk. High blood pressure is the leading risk factor for non-communicable disease mortality in both rich and developing countries. Medical and epidemiological studies have linked ambient air pollution to increased hypertension and cardiovascular stress. One concern in our setting is that agricultural fires may be correlated with unobserved economic conditions that also influence health outcomes. We address this concern with daily data on wind direction, which we use to construct location-specific exposure measures to upwind and downwind fires. The rationale for doing

Non-Technical Summary

so is that upwind fires are likely to affect health purely through changes in pollution levels. We control for downwind fires in our empirical models, thereby capturing any local economic factors associated with fires. The combination of fire activity and wind direction provides a quasi-random source of variation in short-term air pollution exposure.

We leverage data from various sources for this study. To measure fire activity, we use satellite data from NASA's Visible Infrared Imaging Radiometer Suite (VIIRS) Active Fire product. Daily wind direction measures were obtained from the ERA-5 climate reanalysis data. These are all matched with data on blood pressure tests for nearly 784,000 individuals from the National Family and Health Survey (NFHS) – IV. In order to overcome the limitation posed by lack of ground monitoring data, we use satellite and model-derived pollution estimates from the Modern-Era Retrospective analysis for Research and Applications, Version 2. We observe daily location-specific pollution levels for fine particulate matter ($PM_{2.5}$) and its components. Collectively, these data allow us to estimate micro-level reduced form regression specifications for the effects of upwind fires on hypertension risk and air pollution exposure.

Our estimates suggest that, on average, a standard deviation increase in exposure to upwind fires increases the incidence of hypertension by 1.8%. Based on estimates of the value of statistical life (VSL) for India, we find that these fires are associated with monetary costs of the order of $9 billion annually, due to hypertension-related additional mortality. We further show that upwind fires affect $PM_{2.5}$ concentrations, while downwind fires do not. We cannot rule out that other pollutants, also carried by wind and generated from biomass burning, may be partly driving our estimates on hypertension. Nevertheless, our results point to air pollution, collectively, being the primary mechanism through which the fires are affecting health outcomes.

Finally, we also look at heterogeneous effects of these fires by employing the Sorted Effects Method and Classification Analysis. This allows us to identify individual characteristics that are associated with vulnerability to air pollution. Results are well-aligned with epidemiological and medical literature. For example, we find that the estimated effects are significantly stronger for older males, smokers, individuals who were already on blood pressure medication, and individuals belonging to socially marginalized groups.

Air Pollution from Agricultural Fires Increases Hypertension Risk

Hemant K. Pullabhotla* and Mateus Souza‡

Abstract

In many parts of the developing world, farmers widely use deliberate fires to burn vegetation and clear land to plant crops. These agricultural fires, however, are known to be associated with health costs due to increased air pollution. We contribute to underpinning the associated health cost estimates by studying the effects of these fires on hypertension risk. Despite being one of the leading causes of mortality globally, there is little direct evidence on how hypertension risk changes with exposure to pollution from agricultural fires. To overcome common data and empirical challenges in this setting, we match blood pressure readings from nearly 784,000 individuals across India with satellite data on 1.2 million agricultural fires, wind direction realizations, and local ambient air pollution. We find that the incidence of hypertension increases by 1.8% for each standard deviation increase in the number of upwind fires observed one day before the blood pressure readings. We find that the impact is stronger among older males, smokers, individuals that were already on blood pressure medication, and individuals belonging to socially marginalized groups. Our estimates suggest that agricultural fires in India lead to hypertension-related additional mortality, associated with USD 9 billion annually in costs.

Key words: air pollution, agricultural fires, cardiovascular health, hypertension

---

*Center on Food Security and The Environment (FSE), Stanford University and Department of Economics, Deakin University; email: h.pullabhotla@deakin.edu.au.

‡Corresponding Author; Department of Economics, Universidad Carlos III de Madrid; Calle Madrid, 126, Getafe, Madrid, Spain - 28903; email: mateus.nogueira@uc3m.es.

We thank seminar participants at the 2019 AERE Summer Meetings, 2019 AAEA Annual Meeting, 2020 AGU Fall Meeting, FSE seminar at Stanford University, and members of ECHO Lab and Lobell Lab, Stanford University, for their excellent feedback. Souza acknowledges generous support from the European Research Council (ERC, under the European Union's Horizon 2020 research and innovation programme, Grant Agreement No. 772331).

Air Pollution from Agricultural Fires Increases Hypertension Risk

1 Introduction

The global costs of ambient air pollution amount to $6.43 trillion annually, according to recent estimates (World Bank, 2022). Underpinning these estimates are two key components: the value of a statistical life (VSL) and the magnitude of the impact of air pollution on health. Focusing on the latter, an extensive body of work has found adverse effects of air pollution, such as the increased risk of mortality amongst children and adults,[1] and reduced birth weights (Yang and Chou, 2018; Jones and Berrens, 2021). However, it is unclear whether estimates from these prior studies are applicable globally, across diverse contexts and different sources of pollution. For example, recent meta-analyses reveal that most of the existing studies are confined to high-income countries, with little coverage of South Asia, Africa, or Latin America (Orellano et al., 2020; Yang, Qian, et al., 2018; Choi et al., 2019). Extrapolating the concentration-response relations found in developed countries to low-income regions may be problematic given the differences in access to healthcare, dietary and occupational patterns, as well as the much higher levels of ambient air pollution in the developing world (Arceo, Hanna, and Oliva, 2016). Epidemiological evidence also suggests that health damages vary by air pollution source and composition (World Bank, 2022; Brook et al., 2010). However, most empirical studies focus on emissions from urban or industrial sources, with limited attention given to the effects of polluting sources in rural areas. Accurate quantification of the costs of air pollution, and thus the quantification of potential benefits from its abatement, necessarily needs to account for these sources of heterogeneity.

A key challenge limiting the coverage of past studies has been the paucity of air pollution and health data in developing countries, particularly in rural areas. Coverage of air pollution monitors is sparse in many developing countries and limited to large cities or close to industrial areas (Graff Zivin and Neidell, 2013). As a result, evidence on the impact of significant non-urban sources of pollution such as biomass burning, common across rural areas, is minimal. A second empirical challenge, not unique to developing countries, is finding adequate methods to deal with confounding factors. Exposure to air pollution is likely correlated with various unobserved socio-economic dimensions, behavioral patterns, or other environmental factors that also affect health outcomes. Not accounting for such unobserved factors would bias the estimate of the impact of air pollution on health.

[1] For example, see: Chay and Greenstone (2003); Currie et al. (2014); Deryugina et al. (2019); Wu et al. (2020).

In this study, we overcome these data and empirical challenges to provide robust evidence quantifying the health impacts of short-term air pollution exposure from rural emissions in the context of a developing country. We use novel high-resolution satellite data on agricultural fires to generate plausibly exogenous variation in exposure to pollution across India. Like many other parts of the developing world, farmers in India widely use deliberate fires to burn vegetation and clear land to plant crops (Singh and Kaskaoutis, 2014).[2] During peak fire seasons in India, these agricultural fires can contribute to more than half of the particulate pollution load even within urban areas (Liu et al., 2018; Cusworth et al., 2018). It is also important to note that these agricultural fires are controlled burning activities limited to farmers' plots. Unlike wildfires or bushfires, these agricultural fires do not spread uncontrollably and are unlikely to cause damage to property or human life directly. Instead, air pollution from these fires is the primary mechanism through which they are likely to affect health.

We focus on identifying the effects of these fires on hypertension risk. High blood pressure is the leading risk factor for non-communicable disease mortality in both rich and developing countries (Institute for Health Metrics and Evaluation, 2019). Medical and epidemiological studies have linked ambient air pollution to increased hypertension and cardiovascular stress (e.g., Hadley, Vedanthan, and Fuster, 2018; Cosselman, Navas-Acien, and Kaufman, 2015). This body of work has found that small pollution particles can enter the lungs' alveoli, which can trigger systemic inflammation and vasoconstriction, for example. These conditions can increase the risk of heart failure, arrhythmia, and cardiac arrest, among others. However, there are no prior studies that quantify, within a causal framework, the extent to which deliberate biomass burning events are detrimental to exposed individuals' cardiovascular health. One concern is that agricultural fires may be correlated with unobserved economic conditions that also influence health outcomes. We address this concern with daily data on wind direction, which we use to construct location-specific exposure measures to upwind and downwind fires.[3] The rationale for doing so is that upwind fires are likely to affect health purely through changes in pollution levels. We control for downwind fires in our empirical models, thereby capturing any local economic factors associated with fires. The combination of fire activity and wind direction provides a quasi-random source of variation in short-term air pollution exposure.[4]

We leverage data from various sources for this study. To measure fire activity, we use satellite data from

[2]The widespread practice of stubble-burning may be partly attributed to intensification of cropping frequency. Given nonconvexities in net returns to farming, more intensive agricultural practices may have been adopted in response to the pressures from population growth and increased demand for food (Krautkraemer, 1994).

[3]A schematic for our definition of upwind fires is presented in Appendix Figure A.1.

[4]The empirical strategy that we use is similar in spirit to that used in recent studies to examine the impact of pollution from agricultural fires on birth outcomes in Brazil and the effect on test scores among students in China (Rangel and Vogl, 2019; Graff Zivin, Liu, et al., 2020).

NASA's Visible Infrared Imaging Radiometer Suite (VIIRS) Active Fire product (EOSDIS, 2016). With a pixel resolution of 375 meters, we observe more than 1.2 million fire events during our sample period 2015 to 2016. Daily wind direction measures were obtained from the ERA-5 climate reanalysis data (Hersbach et al., 2020). These are all matched with data on blood pressure tests for nearly 784,000 individuals from the National Family and Health Survey (NFHS) – IV (IIPS and ICF, 2017).[5] In order to overcome the limitation posed by lack of ground monitoring data, we use satellite and model-derived pollution estimates from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2; Gelaro et al., 2017), with a grid resolution of 0.625×0.5 degrees.[6] We observe daily location-specific pollution levels for fine particulate matter ($PM_{2.5}$) and its components: Black Carbon, Organic Carbon, SO_2, dust, and sea salt. Collectively, these data allow us to estimate micro-level reduced form regression specifications for the effects of upwind fires on hypertension risk and air pollution exposure.[7]

Our estimates suggest that, on average, a standard deviation increase in exposure to upwind fires increases the incidence of hypertension by 1.8%. Based on estimates of the value of statistical life (VSL) for India (Viscusi and Masterman, 2017), we find that these fires are associated with monetary costs of the order of $9 billion annually, due to hypertension-related additional mortality. We further show that upwind fires affect $PM_{2.5}$ concentrations, while downwind fires do not. We cannot rule out that other pollutants, also carried by wind and generated from biomass burning, may be partly driving our estimates on hypertension. Nevertheless, our results point to air pollution, collectively, being the primary mechanism through which the fires are affecting health outcomes.

Finally, we also look at heterogeneous effects of these fires by employing the Sorted Effects Method and Classification Analysis from Chernozhukov, Fernández-Val, and Luo (2018). This allows us to identify individual characteristics that are associated with vulnerability to air pollution. Results are well-aligned with epidemiological and medical literature. For example, we find that the estimated effects are significantly stronger for older males, smokers, individuals who were already on blood pressure medication, and individuals belonging to socially marginalized groups.

This paper contributes to a growing literature that aims to quantify the health impacts of air pollution.

[5] The NFHS also collects data on blood hemoglobin levels (biomarkers for anemia) and glucose levels (biomarkers for diabetes). However, epidemiological literature suggests that these health outcomes are more likely to be affected by medium- or long-term pollution exposure (e.g., Honda et al., 2017; Lucht et al., 2018). We focus on hypertension, which has been more robustly linked to short-term exposure.

[6] This corresponds to a grid of approximately 50×50 km, depending on the location of the measurement.

[7] We estimate separate regressions for the effects of upwind fires on hypertension risk and air pollution. Following prior literature (Graff Zivin and Neidell, 2013), in this setting we caution against an instrumental variable approach directly linking air pollution and hypertension. The reason is that we do not observe all pollutants associated with the fires, such that the IV approach would violate the exclusion restriction. More details in section 2.2.

While previous studies have often focused on respiratory health as a key physiological mechanism through which air pollution risk is manifested, our results highlight the importance of examining hypertension as an additional pathway of impact. The impact of short-term air pollution exposure on hypertension may partly explain why previous studies find adverse effects also on labor productivity, cognition, mortality, and other crucial human capital outcomes (Zhang, Chen, and Zhang, 2018; Chang et al., 2016; Deryugina et al., 2019). Our findings speak to the potential impacts of other biomass fire events, such as large wildfires, to the extent that they similarly increase air pollution.[8] Our empirical approach could be adapted to examine the health impacts of exposure to air pollution in these settings as well.

2 Main Effects of Agricultural Fires on Hypertension Risk

Figure 1 illustrates the geographic variation in the survey locations of the individual blood pressure tests and three key variables of this study: fire activity; incidence of hypertension; and air pollution. We note that agricultural fires are particularly prevalent in the north-western, north-eastern, and eastern regions of the country. These are districts where farmers either use fire to clear their harvest residue on their fields between two cropping seasons or use land-clearing fires in districts where shift and burn agriculture is practiced (Singh and Kaskaoutis, 2014). Across the full sample, individuals are exposed to 2.8 upwind fires, on average, in the 24 hour period leading to the day of the health tests. Considering only exposed individuals, the average number of upwind fires is closer to 9.6.

2.1 Main Empirical Specification and Results

We implement reduced form regression analyses to formally examine the causal link between agricultural fire activity and cardiovascular distress. For identification of that link, we construct a variable that counts the number of "upwind fires" to which individuals in our sample are exposed. We count the number of fire pixels measured using the Visible Infrared Imaging Radiometer Suite (VIIRS) 375 m thermal anomalies and active fires data (EOSDIS, 2016) within a buffer of radius 50, 75, 100 or 150 km around each individual's location in our sample. Those fire counts are then classified as located upwind, downwind, or other directions, based on wind direction data from ERA-5 (Hersbach et al., 2020). This classification is presented in further detail in the schematic in Appendix Figure A.1. The exposure variable of interest is the number of upwind fires on

[8] Note, however, that different types of biomass burning events are significantly heterogeneous with regards to the pollutants that they generate, and how those are dispersed in the atmosphere (Andreae, 2019; Vicente et al., 2013).

the day before the blood pressure tests.[9] Our identification strategy thus relies on combining variations in wind directions and fire activity, generating plausibly exogenous variation in exposure to agricultural fires.[10]

Our regressions also incorporate district-by-month-of-sample fixed effects, to flexibly account for district and season-specific factors that vary across space. Effectively, our specifications compare blood pressure outcomes for individuals located within the same district and measured within the same month but exposed to different quantities of upwind fires on the day of the blood pressure test. On average, each month of the survey covers 33,000 individuals across 55 districts. The survey's roll-out determines the specific day within a district-month group on which an individual's blood pressure is measured. Therefore, the measurement date is unlikely to be correlated with any individual or household factors that affect health. This variation in the day of blood pressure measurement, combined with the plausibly exogenous variation in wind directions, allows us to estimate the causal effect of agricultural fire exposure on hypertension. In addition to the fixed effects, all of our specifications include: the number of fires in downwind and other (non up/down) directions; weather variables such as rainfall, temperature, wind direction, and wind speed; as well as a rich set of individual and household characteristics. These account for any local economic, agricultural or other factors which might be correlated with agricultural fires, pollution, and health outcomes. The exact regression specification is presented in Appendix B.1. Descriptive statistics for the outcomes and for the control variables are presented in Appendix Tables A.1 and A.2, respectively.

The main outcome of interest is a binary indicator variable for incidence of hypertension (i.e., systolic blood pressure $\geq$ 140 mmHg or diastolic blood pressure $\geq$ 90 mmHg).[11] We multiply this variable by one thousand, such that the coefficients can be interpreted as marginal effects on incidence per thousand ($'000$).[12] Figure 2 presents results from our main specifications. We find that upwind fires increase hypertension risk, while fires located downwind do not. Particularly, for a specification looking at fires within a 75km radius, each upwind fire increases incidence of high blood pressure by 0.12 per thousand. For each standard

[9] In Appendix C.4 we show that upwind fires from more than one day before the tests, or from days after the tests, do not significantly affect hypertension risk. We thus focus on the fires on the day immediately before the tests.

[10] We stress that our main source of quasi-exogenous variation are changes in wind directions, and not fire activity itself. In Appendix Figure C.4 we show that there is substantial day-to-day variation in wind directions during the NFHS survey period. In about 33% of the cases, wind direction quadrants on the interview day were different from those on the day before. This share is even higher (48%) if we define wind direction based on octants. The implication is that farmers are unlikely to be able to predict day-to-day changes in wind direction, thus have little scope to manipulate our main exposure variable (upwind fires).

[11] Systolic BP measures arterial pressure when the heart beats, while diastolic BP measures arterial pressure when the heart rests. We define hypertension based on the NFHS cutoffs for "Abnormal (Mildly Elevated)" blood pressure (IIPS and ICF, 2017). We focus on this binary outcome because clinical interventions are only recommended when BP crosses given thresholds (Messerli, Williams, and Ritz, 2007). An increase in BP, per se, does not pose a threat to health if BP still remains at normal levels.

[12] Per thousand, or '000, is typically used in medical literature to measure the incidence of given health conditions, such as hypertension. It represents how many people, out of one thousand, suffer from the condition. This is a normalized measure, much like percentages, which allows us to compare incidence across samples or over time.

deviation increase in upwind fires (approximately 14 fires), the incidence of high blood pressure increases 1.68 per thousand, representing about 1.8% of the average incidence for individuals that were not exposed. Figure 2 also shows that results remain similar with alternative radii.[13] Regression estimates corresponding to Figure 2 are presented in Appendix Table C.1.

Our estimates remain robust across a variety of model specifications and robustness checks. We obtain similar results using a logit estimation (Appendix Table C.2).[14] In Appendix Table C.3 we show that the above results hold with a specification that does not include weather or demographic controls, suggesting that the district-by-month-of-sample fixed effects may already capture most of the confounding variation in this context. The effects that we find are also robust to using more demanding sets of fixed effects. In Appendix Table C.4 we see that the coefficient on upwind fires remains qualitatively similar as we progressively change the fixed effects from district-by-month-of-sample to NFHS sample cluster-by-week-of-sample.[15] Results are also statistically significant with an alternative definition of hypertension risk, as shown in Table C.5. Although, in that case, the magnitudes of the effects seem smaller compared to the sample average incidence of hypertension. Table C.6 shows that results hold if we classify upwind and downwind fires based on 45-degree octants (rather than 90-degree quadrants used in the main specification).

Finally, in Appendix C.5 we consider variants of our main specification with continuous blood pressure measures (systolic and diastolic) as the outcomes. Results are shown in Appendix Figure C.2. We show that the coefficients on downwind fires are null across all specifications. Conversely, we find positive and statistically significant effects for upwind fires, but only for subsamples of "high risk" individuals who had systolic BP $\geq$ 120; diastolic BP $\geq$ 80. Additionally, heterogeneity analyses in Appendix Figure C.3 suggest stronger effects for individuals with body-mass index (BMI) $\geq$ 30, or with age $\geq$ 40.[16] Collectively, these

[13] As one would expect, increasing the distance used to measure exposure results in smaller marginal effects on blood pressure. This reduction in impact is consistent with the air quality effect of fires reducing with distance. Conversely, shorter distances limit the available variation in the number of fires we detect in our data, leading to a loss of precision. We also run into the issue of measurement error in the individual's location with shorter distances – the geocoordinates of the sample cluster locations in the NFHS data are randomly displaced by up to 10-kilometers in rural areas for preserving the privacy of respondents (Burgert et al., 2013). Finally, the climate reanalysis data that we use is too coarse to calculate reliably define up and down-wind directions at smaller distances increasing the potential for misclassifying up and down-wind fires. Therefore, we focus on the results using a 75-kilometer radius buffer in our main discussion.

[14] Our findings are consistent with those from prior studies in similar settings. When converted to odds ratios, our estimates suggest that a one s.d. increase in upwind fire exposure (about 14 fires) increases the odds of hypertension risk by 1.02. Recent estimates from California suggest that one day of wildfire smoke exposure increases the relative risk of emergency department visits for hypertension by between 1.01 to 1.08 for light, medium, and dense smoke, respectively (Wettstein et al., 2018). Similarly, Singh, Roy, et al. (2021) find that high intensity biomass burning events are associated with odd ratios ranging from 1.003 to 1.32, for hypertension among populations in four North Indian states.

[15] NFHS sample clusters are survey enumeration areas roughly corresponding to a village in rural areas. This is the lowest level at which geographic location information is available in the NFHS.

[16] Coefficients from Appendix Figure C.3 were obtained from the interactions of upwind fires with given categorical variables of interest.

results suggest that agricultural fire-induced hypertension risk may be particularly prevalent for a subset of vulnerable individuals. Estimates of average effects across a broad sample may mask this risk. In section 3 we perform a deeper analysis of heterogeneity, to provide further insight about risk factors.

2.2 Air Pollution as the Mechanism Linking Fire Activity and Cardiovascular Distress

Based on medical and epidemiological literature,[17] we argue that fire activity increases hypertension risk primarily through increased air pollution. To support this argument, we provide evidence within our study setting that increased fire activity indeed leads to higher pollution concentrations. We use regression specifications similar to those described in section 2.1, but with measures of air pollution as the dependent variables. In particular, we focus on fine particulate matter ($PM_{2.5}$) concentrations. We construct daily average $PM_{2.5}$ measures for each NFHS cluster location (i.e., village) using climate reanalysis data from MERRA-2 (Gelaro et al., 2017).[18] Additionally, we run separate regressions for sulfate (SO_2), organic carbon (OC), black carbon (BC), dust, and sea-salt, which are all components of our $PM_{2.5}$ measure.

Results in Figure 3 and Appendix Table D.1 show that upwind fires within 75km surrounding NFHS respondents' locations significantly increase air pollution concentrations. Consistent with wind transporting pollution from fires, we see that upwind fires have a relatively larger effect compared to downwind fires on Black Carbon, Organic Carbon, and SO_2 – all of which are pollutants commonly associated with biomass fires (Akagi et al., 2011). Dust and sea-salt particulate concentrations, on the other hand, are determined by long-range atmospheric transport and proximity to arid or desert regions and are less likely to be affected by fires. Consequently, we see that the impacts of fires on dust and sea salt particulate matter are close to zero. Conversely, the effects on the combined $PM_{2.5}$ are substantial. We find that a standard deviation increase of exposure to upwind fires increases total $PM_{2.5}$ by about 27%, relative to the unexposed group average of 38.6 $\mu g/m^3$.

The above results use modeled pollution measures from satellite data and atmospheric chemical transport estimates. We find very similar estimates using the limited amount of monitoring station data available in India for the sample period (Appendix Figure D.1). In Appendix Table D.2 and Figure D.2, we see that a

[17] Among others, see: Al-Kindi et al. (2020); Sanidas et al. (2017); Cosselman, Navas-Acien, and Kaufman (2015); Mills et al. (2009); Prabhakaran et al. (2020).

[18] We use the MERRA-2 hourly estimates for surface concentrations of sulfate, organic carbon, black carbon, dust and sea-salt particulate matter with a diameter of less than 2.5 μm. We combine these species level estimates to obtain total $PM_{2.5}$ concentrations using ground-validated conversion factors from He et al. (2019). We aggregate the hourly estimates to 24-hourly averages for use in the analysis.

standard deviation increase in upwind fires increases $PM_{2.5}$ measured at monitors by 26%, relative to the mean $PM_{2.5}$ of 83.8 $\mu g/m^3$ on monitor-days with zero upwind fire – a magnitude very similar to what we find using modeled $PM_{2.5}$ data across the NFHS sample locations. Additionally, we also find that fires affect multiple pollutants simultaneously with upwind fires increasing PM_{10}, NO, and NO_2 concentrations.

The significant associations shown in Figure 3 may lead one to consider using upwind fires as an instrumental variable (IV) to estimate the direct effect of particulate matter exposure on hypertension risk. In Appendix E we present results from one such approach. However, we refrain from interpreting those results as causal, and we caution against an IV strategy in settings similar to ours, for two main reasons. First, evidence from the atmospheric science literature shows that biomass burning events are associated not only with $PM_{2.5}$ emissions, but also with emissions of carbon monoxide, nitrogen oxides, volatile organic compounds (VOCs), and ozone, among others (Andreae, 2019; Vicente et al., 2013). Second, epidemiological studies suggest that these other pollutants may also be associated with hypertension risk (for reviews, see: Yang, Qian, et al., 2018; Orellano et al., 2020). The implication is that upwind fires are not a valid instrument for $PM_{2.5}$, due to violations of the "exclusion restriction," and to the extent that wind also carries other pollutants.[19]

In that case, it is challenging to identify which pollutants are more strongly associated with hypertension risk. This is particularly true in developing country settings, where reliable measurements and comprehensive measures of air pollution are scarce (World Bank, 2022). In Appendix E we show that, for our setting, two-stage least squares (2SLS) estimates of the effects of $PM_{2.5}$ on hypertension risk are sensitive to violations of the exclusion restriction. We perform simulations based on Conley, Hansen, and Rossi (2012), to adjust the 2SLS confidence intervals, assuming that other pollutants also affect hypertension. Results suggest that the 2SLS estimates are no longer consistent if there exists another pollutant with at least one third of the emissions factor of $PM_{2.5}$, and which similarly causes at least one third of the health damages.

We stress that this does not imply that air pollution, as an ensemble of pollutants, may not be the main mechanism through which upwind agricultural fires affect hypertension. Recall that our main specifications in section 2.1 control for fires in other wind directions, which we find to be associated with null and non-significant coefficients. The implication is that whatever is causing an increase in hypertension risk must be carried, by wind, from the locations where the fires are happening, in direction to the surveyed villages. The natural conclusion is that air pollution is the primary mechanism in this context.

[19] Our arguments are in line with Graff Zivin and Neidell (2013), who stress the complexity of identifying the effects of multiple pollutants in these types of settings. Rather, they recommend estimating reduced form specifications linking the outcomes of interest directly to the source of pollution (i.e., agricultural fires).

3 Heterogeneity and Characteristics of Most Affected Individuals

To identify the heterogeneity of the effects of agricultural fires across individuals in our sample, we implement the Sorted Effects Method (SEM) and Classification Analysis (CA) (Chernozhukov, Fernández-Val, and Luo, 2018). SEM consists of rerunning the analysis from section 2.1 but at multiple quantiles of the available data (further details in Appendix B.2). Specifically, we implement the SEM for a binary response model (predicting the incidence of hypertension), a continuous independent variable of interest (number of upwind fires), plus the controls described in section 2. With SEM, we identify the partial effect of the upwind fires, holding all other factors constant.

Figure 4 presents the main results from the SEM, where we rank the estimated partial effects from lowest to highest. This ranking identifies portions of the sample distribution likely to be least or most affected by the fires. The estimates reveal significant heterogeneity across the sample. The least affected group (bottom 10th percentile) experience increases of high blood pressure incidence of 0.02 per thousand for exposure to each additional upwind fire. On the other hand, in the most affected group, the effects can be as large as 0.29 per thousand (almost fifteen times larger). Once the least and most affected groups are identified, it is possible to compare their characteristics with a Classification Analysis (CA; further details presented in Appendix B.2).

Table 1 presents results from the Classification Analysis. We calculate averages for key variables of interest across groups identified as least and most affected by upwind fires. We then take the differences in averages. Bootstrap-based inference techniques are used to test for statistically significant differences in these characteristics across the two groups. We find that variation in upwind fire's impact along individual physiological and behavioral characteristics is consistent with known risk factors for hypertension from the medical literature. For instance, we see that 18.3% of the most affected individuals already held prescriptions for blood pressure medication. Also, 5% of them were smokers. Virtually no individuals from the least affected group were smokers or reported having blood pressure prescriptions. As expected, age and BMI are important signals of vulnerabilities. Most affected individuals are, on average, 43 years old, while the least affected are younger than 19. The average BMI of most affected individuals is 26.5, above the "overweight" cutoff (WHO Expert Consultation, 2004). Conversely, the average BMI of the least affected individuals is 18.4, closer to normal ranges.

Our analysis also reveals significant heterogeneity in the impact of fires on hypertension across gender. Almost 30% of the most affected individuals were male, compared to only 2.5% of the least affected group

(here we note that our total sample consists of 24% males). Differences in biological mechanisms across gender, resulting in a higher risk of hypertension among men (Song et al., 2020), may be partly responsible for this gender differential. Additionally, in the study setting, men may also be more likely to be exposed to pollution from agricultural fires due to differences in gender roles and occupational structure. In Indian rural communities, men often work in agriculture or other outdoor activities, while women traditionally focus on in-home activities. We also see that social marginalization is another critical dimension that can exacerbate air pollution's health impacts. Our heterogeneity results suggest that the most affected individuals have a higher likelihood of belonging to Scheduled Tribes – a group that was historically marginalized.

We see a decline in the risk of hypertension due to air pollution with improved education levels. Most affected individuals are typically less educated (with an average of 5.8 years), even though they are older. This variation may partly reflect advances in the Indian education system and living standards, where younger generations attain more schooling and are less likely to work in agriculture. It could also be that more educated individuals engage in defensive behaviors, such as wearing masks or staying indoors during high pollution events. However, our results suggest that even though some households may be able to invest in technologies to improve the air quality within their homes, they cannot fully mitigate the detrimental health effects of ambient air pollution. For instance, we find that individuals with higher risk are more likely to belong to households that cook outdoors and are more likely to use clean cooking fuels, which would reduce indoor air pollution levels. We also see that the most affected individuals are likely to belong to wealthier households. These results highlight the importance of addressing outdoor air pollution in this setting.

4 Conclusion

We provide causal estimates of the extent to which agricultural fires increase hypertension risk. Our empirical strategy combines micro-level data on fire activity and wind directions to generate quasi-random variation in air pollution exposure. Our results show that a standard deviation increase of exposure to upwind fires increases the incidence of hypertension by 1.8% – relative to the average hypertension incidence among individuals not exposed to upwind fires. With a heterogeneity analysis, we also show that the effects can be significantly larger for more vulnerable individuals. For example, we find that older males, smokers, individuals on blood pressure medication, and individuals belonging to socially marginalized groups are significantly more vulnerable.

Using the nationally representative nature of our data, we estimate that the exposure to fires contributes to

nearly 14% of total mortality due to hypertension in India annually. Using value of statistical life estimates, we find that the additional mortality imposes a substantial monetary cost of USD 9 billion each year (2.29 billion lower bound, 18.24 billion upper bound).[20]

The large health burden posed by fires reflects the fact that agricultural fires are a major contributor to overall pollution in India. Overall, agricultural fires are estimated to contribute to nearly a quarter of all black carbon, organic carbon, and carbon monoxide emissions in India (Venkataraman et al., 2006). In peak seasons, fire can contribute to nearly half of particulate pollution even in urban areas such as the capital city of Delhi (Liu et al., 2018). We also find that exposure to fires is widespread. Within our sample, nearly a third of respondents were exposed to at least one fire in the day before BP tests and, on average, experienced almost 38 days with upwind fires within a year. Finally, overall ambient air pollution is estimated to be the leading reason accounting for 22-53% of all deaths from cardiovascular diseases (Nair et al., 2021).

While previous studies have often focused on respiratory health as a key physiological mechanism through which air pollution risk is manifested, our results highlight the importance of examining hypertension as an additional pathway of impact. The impact of short-term air pollution exposure on hypertension may explain why previous studies find adverse effects also on labor productivity, cognition, mortality and other important human capital outcomes (Zhang, Chen, and Zhang, 2018; Chang et al., 2016; Deryugina et al., 2019). Our findings also speak to the potential adverse health impacts of other biomass fire events, such as large wildfires, which are also prevalent in some developed economies.

[20]Details of these calculations are presented in Table B.1.

References

Akagi, SK et al. (2011). "Emission factors for open and domestic biomass burning for use in atmospheric models". *Atmospheric Chemistry and Physics* 11, 4039–4072.

Andreae, M. O. (2019). "Emission of trace gases and aerosols from biomass burning – an updated assessment". *Atmospheric Chemistry and Physics* 19, 8523–8546.

Andrews, Isaiah, James H. Stock, and Liyang Sun (Aug. 2019). "Weak Instruments in Instrumental Variables Regression: Theory and Practice". *Annual Review of Economics* 11, 727–753.

Arceo, Eva, Rema Hanna, and Paulina Oliva (Jan. 2016). "Does the Effect of Pollution on Infant Mortality Differ Between Developing and Developed Countries? Evidence from Mexico City". *The Economic Journal* 126, 257–280.

Brook, Robert D et al. (2010). "Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association". *Circulation* 121, 2331–2378.

Burgert, Clara R, Josh Colston, Thea Roy, and Blake Zachary (2013). *Geographic displacement procedure and georeferenced data release policy for the Demographic and Health Surveys*. ICF International.

Chang, Tom, Joshua Graff Zivin, Tal Gross, and Matthew Neidell (2016). "Particulate Pollution and the Productivity of Pear Packers". *American Economic Journal: Economic Policy* 8, 141–69.

Chay, Kenneth Y. and Michael Greenstone (Aug. 2003). "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession". *The Quarterly Journal of Economics* 118, 1121–1167.

Chen, Shuowen, Victor Chernozhukov, Ivan Fernández-Val, and Ye Luo (2020). "Sorted Effects: Estimation and Inference Methods for Sorted Causal Effects and Classification Analysis". *R package version 1.2.0*.

Chernozhukov, Victor, Iván Fernández-Val, and Ye Luo (2018). "The Sorted Effects Method: Discovering Heterogeneous Effects Beyond Their Averages". *Econometrica* 86, 1911–1938.

Choi, You-Jung et al. (Dec. 2019). "Short-term effects of air pollution on blood pressure". *Scientific Reports* 9, 20298.

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi (2012). "Plausibly Exogenous". *The Review of Economics and Statistics* 94, 260–272.

Cosselman, Kristen E., Ana Navas-Acien, and Joel D. Kaufman (Nov. 2015). "Environmental factors in cardiovascular disease". *Nature Reviews Cardiology* 12, 627–642.

Currie, Janet, Joshua Graff Zivin, Jamie Mullins, and Matthew Neidell (2014). "What Do We Know About Short- and Long-Term Effects of Early-Life Exposure to Pollution?" *Annual Review of Resource Economics* 6, 217–247.

Cusworth, Daniel H et al. (Apr. 2018). "Quantifying the influence of agricultural fires in northwest India on urban air pollution in Delhi, India". *Environmental Research Letters* 13, 044018.

Dai, Yue et al. (2019). "Short-term and long-term blood pressure changes and the risk of all-cause and cardiovascular mortality". *BioMed Research International* 2019.

Deryugina, Tatyana, Garth Heutel, Nolan H. Miller, David Molitor, and Julian Reif (Dec. 2019). "The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction". *American Economic Review* 109, 4178–4219.

EOSDIS (2016). "VIIRS active fire locations 375m FIRMS V001 NRT". *NRT VIIRS 375m Active Fire Product VNP14IMGT*.

Gelaro, Ronald et al. (2017). "The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)". *Journal of Climate* 30, 5419–5454.

Global Burden of Disease Collaborative Network (2020a). *Global Burden of Disease Study 2019 (GBD 2019) Population Estimates 1950-2019*. Institute for Health Metrics and Evaluation (IHME).

Global Burden of Disease Collaborative Network (2020b). *Global Burden of Disease Study 2019 (GBD 2019) Results*. Institute for Health Metrics and Evaluation (IHME).

Graff Zivin, Joshua, Tong Liu, Yingquan Song, Qu Tang, and Peng Zhang (Nov. 2020). "The unintended impacts of agricultural fires: Human capital in China". *Journal of Development Economics* 147, 102560.

Graff Zivin, Joshua and Matthew Neidell (2013). "Environment, health, and human capital". *Journal of Economic Literature* 51, 689–730.

Hadley, Michael B., Rajesh Vedanthan, and Valentin Fuster (Apr. 2018). "Air pollution and cardiovascular disease: a window of opportunity". *Nature Reviews Cardiology* 15, 193–194.

He, Lijie, Aiwen Lin, Xinxin Chen, Hao Zhou, Zhigao Zhou, and Peipei He (2019). "Assessment of MERRA-2 surface $PM_{2.5}$ over the Yangtze river Basin: Ground-based verification, spatiotemporal distribution and meteorological dependence". *Remote Sensing* 11, 460.

Hersbach, Hans et al. (July 2020). "The ERA5 global reanalysis". *Quarterly Journal of the Royal Meteorological Society* 146, 1999–2049.

Honda, Trenton, Vivian C. Pun, Justin Manjourides, and Helen Suh (2017). "Anemia prevalence and hemoglobin levels are associated with long-term exposure to air pollution in an older population". *Environment International* 101, 125–132.

Institute for Health Metrics and Evaluation (2019). *GBD Compare — IHME Viz Hub — Global*.

International Institute for Population Sciences (IIPS), and ICF (2017). "National Family and Health Survey (NFHS-4), 2015-16".

Jones, Benjamin A. and Robert P. Berrens (2021). "Prescribed Burns, Smoke Exposure, and Infant Health". *Contemporary Economic Policy* 39, 292–309.

Al-Kindi, Sadeer G., Robert D. Brook, Shyam Biswal, and Sanjay Rajagopalan (Oct. 2020). "Environmental determinants of cardiovascular disease: lessons learned from air pollution". *Nature Reviews Cardiology* 17, 656–672.

Krautkraemer, Jeffrey A. (1994). "Population growth, soil fertility, and agricultural intensification". *Journal of Development Economics* 44, 403–428.

Lee, David S, Justin McCrary, Marcelo J Moreira, and Jack R Porter (2021). "Valid t-ratio Inference for IV". *NBER Working Paper* 29124.

Liu, Tianjia et al. (2018). "Seasonal impact of regional outdoor biomass burning on air pollution in three Indian cities: Delhi, Bengaluru, and Pune". *Atmospheric Environment* 172, 83–92.

Lucht, Sarah A et al. (2018). "Air Pollution and Glucose Metabolism: An Analysis in Non-Diabetic Participants of the Heinz Nixdorf Recall Study". *Environmental Health Perspectives* 126.

Messerli, Franz H, Bryan Williams, and Eberhard Ritz (2007). "Essential hypertension". *The Lancet* 370, 591–603.

Mills, Nicholas L. et al. (Jan. 2009). "Adverse cardiovascular effects of air pollution". *Nature Clinical Practice Cardiovascular Medicine* 6, 36–44.

Nair, Moorthy, Hemant Bherwani, Shahid Mirza, Saima Anjum, and Rakesh Kumar (2021). "Valuing burden of premature mortality attributable to air pollution in major million-plus non-attainment cities of India". *Scientific reports* 11, 1–15.

Orellano, Pablo, Julieta Reynoso, Nancy Quaranta, Ariel Bardach, and Agustin Ciapponi (2020). "Short-term exposure to particulate matter (PM10 and PM2.5), nitrogen dioxide (NO2), and ozone (O3) and all-cause and cause-specific mortality: Systematic review and meta-analysis". *Environment International* 142, 105876.

Prabhakaran, Dorairaj et al. (2020). "Exposure to Particulate Matter Is Associated With Elevated Blood Pressure and Incident Hypertension in Urban India". *Hypertension* 76, 1289–1298.

Prenissl, Jonas et al. (May 2019). "Hypertension screening, awareness, treatment, and control in India: A nationally representative cross-sectional study among individuals aged 15 to 49 years". *PLoS Medicine* 16. Ed. by Margaret E. Kruk, e1002801.

Rangel, Marcos A. and Tom S. Vogl (Sept. 2019). "Agricultural Fires and Health at Birth". *The Review of Economics and Statistics* 101, 616–630.

Sanidas, Elias et al. (2017). "Air pollution and arterial hypertension. A new risk factor is in the air". *Journal of the American Society of Hypertension* 11, 709–715.

Singh, Prachi, Ambuj Roy, Dinkar Bhasin, Mudit Kapoor, Shamika Ravi, and Sagnik Dey (2021). "Crop Fires and Cardiovascular Health – A Study from North India". *SSM - Population Health* 14, 100757.

Singh, Ramesh P. and Dimitris G. Kaskaoutis (Sept. 2014). "Crop Residue Burning: A threat to South Asian air quality". *Eos, Transactions American Geophysical Union* 95, 333–334.

Song, Juan Juan, Zheng Ma, Juan Wang, Lin Xi Chen, and Jiu Chang Zhong (Feb. 2020). "Gender differences in hypertension". *Journal of Cardiovascular Translational Research* 13, 47–54.

Venkataraman, Chandra et al. (2006). "Emissions from open biomass burning in India: Integrating the inventory approach with high-resolution Moderate Resolution Imaging Spectroradiometer (MODIS) active-fire and land cover data". *Global biogeochemical cycles* 20.

Vicente, Ana et al. (2013). "Emission factors and detailed chemical composition of smoke particles from the 2010 wildfire season". *Atmospheric Environment* 71, 295–303.

Viscusi, W Kip and Clayton J Masterman (2017). "Income elasticities and global values of a statistical life". *Journal of Benefit-Cost Analysis* 8, 226–250.

Wettstein, Zachary S, Sumi Hoshiko, Jahan Fahimi, Robert J Harrison, Wayne E Cascio, and Ana G Rappold (2018). "Cardiovascular and cerebrovascular emergency department visits associated with wildfire smoke exposure in California in 2015". *Journal of the American Heart Association* 7, e007492.

WHO Expert Consultation (2004). "Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies". *Lancet* 363, 157–63.

World Bank (2022). "The Global Health Cost of PM2.5 Air Pollution: A Case for Action Beyond 2021". *International Development in Focus*.

Wu, X., D. Braun, J. Schwartz, M. A. Kioumourtzoglou, and F. Dominici (2020). "Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly". *Science Advances* 6.

Yang, Muzhe and Shin-Yi Chou (2018). "The impact of environmental regulation on fetal health: Evidence from the shutdown of a coal-fired power plant located upwind of New Jersey". *Journal of Environmental Economics and Management* 90, 269–293.

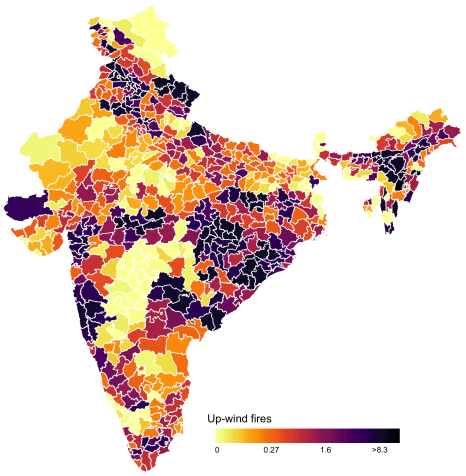
Yang, Bo-Yi, Zhengmin Qian, et al. (2018). "Global association between ambient air pollution and blood pressure: A systematic review and meta-analysis". *Environmental Pollution* 235, 576–588.

Zhang, Xin, Xi Chen, and Xiaobo Zhang (2018). "The impact of exposure to air pollution on cognitive performance". *Proceedings of the National Academy of Sciences* 115, 9193–9197.

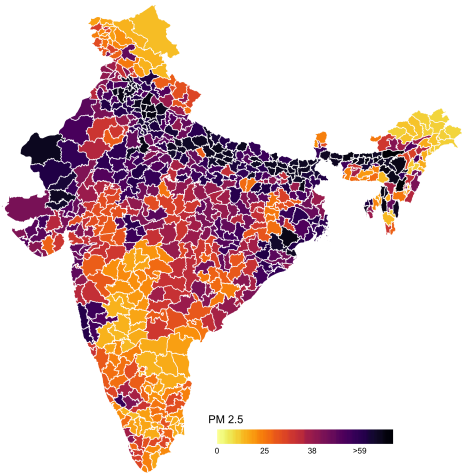
Figures

Figure 1: Distribution of Sample Locations, Fire Activity, Air Pollution, and High Blood Pressure Incidence

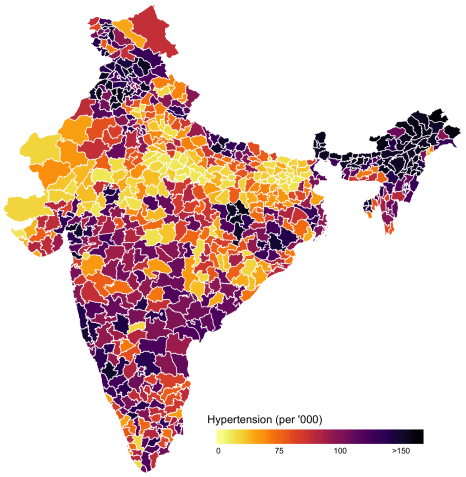
(a) Location of NFHS-IV sample clusters



(b) Fire Activity



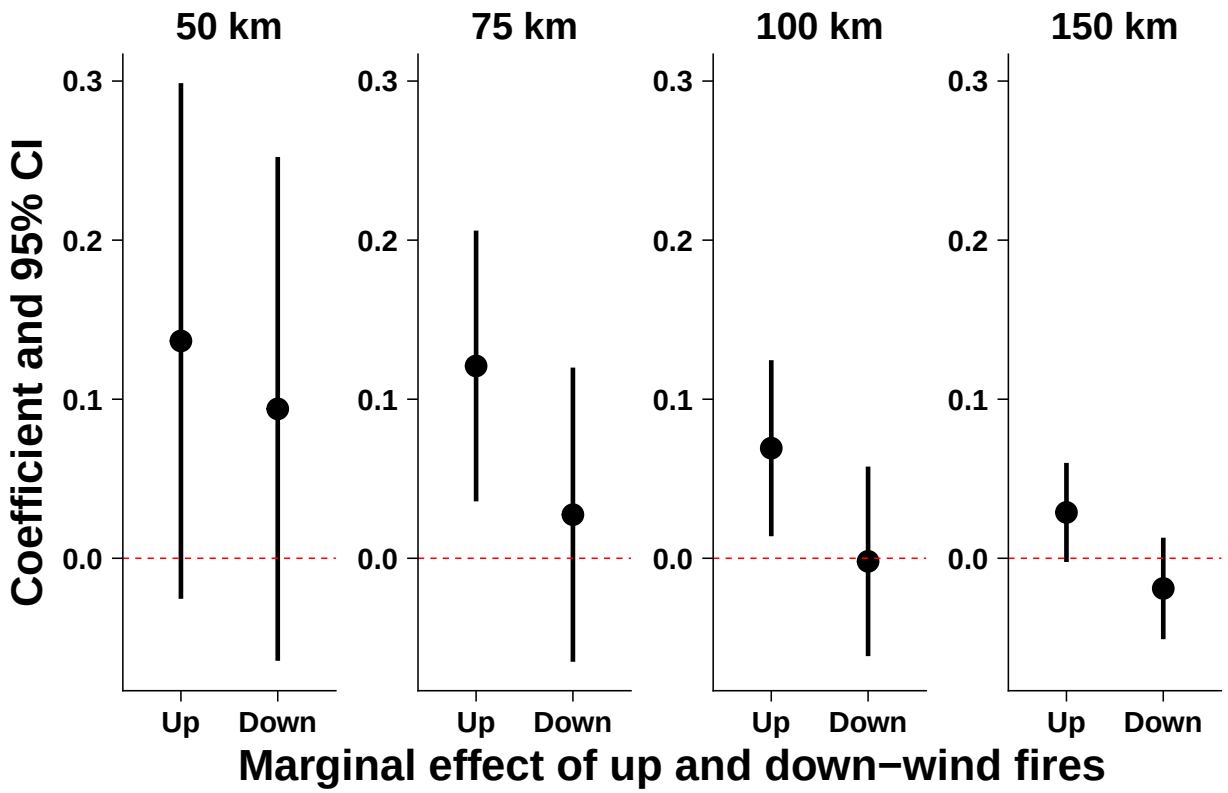
(c) Air Pollution ($PM_{2.5}$)



(d) Incidence of Hypertension

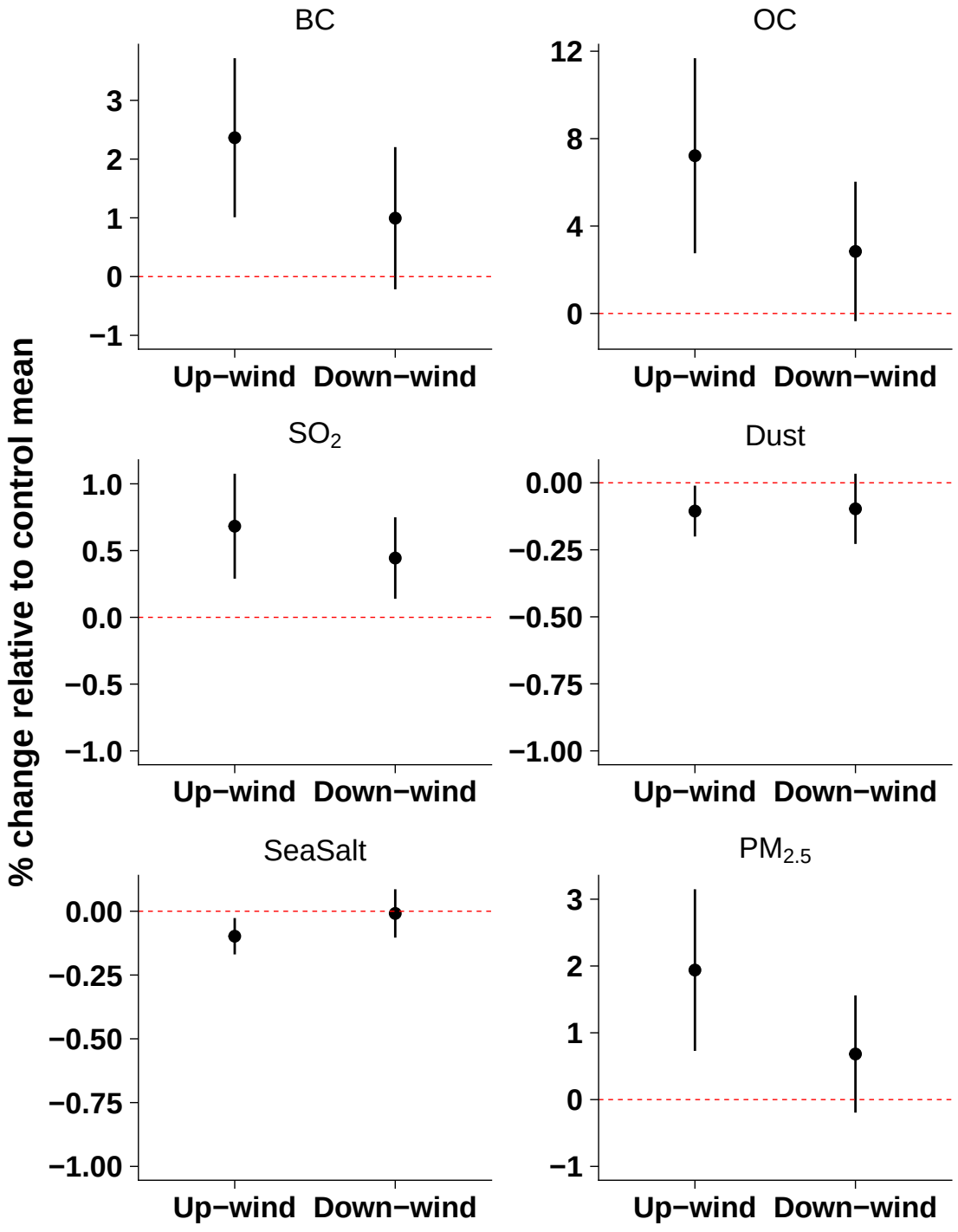
Notes: Panel (a) shows the sample locations (clusters) in the NFHS-IV survey data used in our study. Panel (b) shows the spatial distribution of the number of upwind fires in the 24 hours leading to the day of the blood pressure test, averaged at the district-level. The measure is based on the number of fire pixels from the VIIRS data EOSDIS (2016) observed in a 75-kilometer radius surrounding the survey respondent's location. We classify fires as upwind using data on wind direction from ERA-5 climate reanalysis (Hersbach et al., 2020). "upwind" refers to the direction from which the wind is blowing relative to the location of interest (see Figure A.1). Panel (c) portrays the distribution of average $PM_{2.5}$ exposure at the district level on the day leading to the blood pressure test. The sample mean $PM_{2.5}$ exposure is 39.7 $\mu g/m^3$. Panel (d) shows the mean prevalence of hypertension (systolic blood pressure (BP)$\geq$ 140 mmHg or diastolic BP$\geq$ 90 mmHg) across districts. The sample mean hypertension incidence is 94.7 per $'000$.

Figure 2: Impact of exposure to fires on the incidence of hypertension



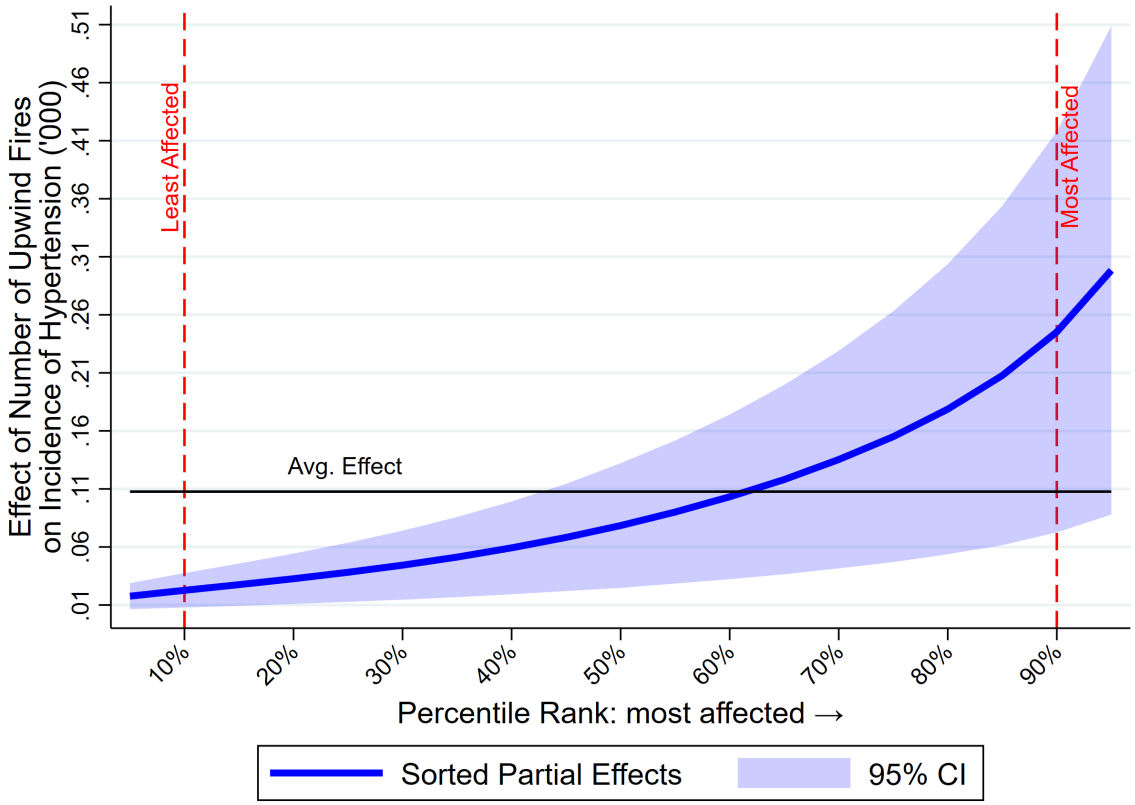
Notes: The figure shows the marginal effect of exposure to one additional fire on the day leading to the blood pressure (BP) test on the risk of hypertension. Coefficients and associated 95% confidence intervals are shown from separate regressions in each column using 50, 75, 100, and 150 km radius buffers for fire counts, respectively. The dependent variable is a binary variable for the incidence of hypertension which takes the value 1000 if the average of three BP tests shows systolic BP $\geq$ 140 mmHg or diastolic BP$\geq$ 90 mmHg, and zero otherwise. Up and down-wind fires classification is based on 90-degree wind sectors. All specifications include district-by-month of sample and day of week fixed effects, controls for weather, and demographic and household characteristics.

Figure 3: Effects of agricultural fires on local air pollution



Notes: Plots show the marginal effects of up- and downwind agricultural fires in a 75-kilometer radius surrounding the survey respondent's location on air pollution concentrations measured using data from atmospheric chemistry transport models (MERRA-2). Coefficients are transformed to show the percentage change relative to the sample mean of the outcome variable along with the associated 95% confidence intervals. Each figure shows estimates from a separate regression for Black Carbon (BC), Organic Carbon (OC), Sulphur dioxide (SO_2), Dust , and Sea Salt particulate matter. We combine these species-level estimates to obtain total $PM_{2.5}$ concentrations using ground-validated conversion factors from He et al. (2019). The effect on $PM_{2.5}$ is shown in the last panel. Up- and downwind is based on 90-degree wind sectors. All specifications control for rainfall, temperature, wind speed, and include district-by-month of sample and day of week fixed effects.

Figure 4: Results from the Sorted Effects Method



Notes: This figure presents the results from the Sorted Effects Method (Chernozhukov, Fernández-Val, and Luo, 2018). The blue line represents the partial effects of upwind fires on incidence of high blood pressure, holding all other factors constant. The blue shaded area represents 95% confidence intervals. The vertical red dashed lines represent the cutoffs for the most (top 10%) and least (bottom 10%) affected individuals.

Tables

Table 1: Results from the Classification Analysis

	(1) Most Affected	(2) Least Affected	(1) - (2) Difference
Prescription for BP (yes=1)	0.1835	0.0000	0.1835
	(0.0018)	(0.0001)	(0.0018)
Smoker (yes=1)	0.0499	0.0014	0.0485
	(0.0016)	(0.0002)	(0.0018)
Male (yes=1)	0.2974	0.0250	0.2723
	(0.0029)	(0.0014)	(0.0036)
Scheduled Tribe (yes=1)	0.2167	0.0890	0.1277
	(0.0026)	(0.0037)	(0.0056)
Age	43.0620	18.6394	24.4225
	(0.0480)	(0.0576)	(0.0793)
Clean Cooking Fuel (yes=1)	0.5147	0.3105	0.2042
	(0.0036)	(0.0063)	(0.0086)
Richest Households (yes=1)	0.2613	0.1598	0.1015
	(0.0032)	(0.0036)	(0.0058)
BMI	26.5612	18.4224	8.1388
	(0.0329)	(0.0275)	(0.0495)
Outdoor Cooking (yes=1)	0.7763	0.6228	0.1534
	(0.0030)	(0.0056)	(0.0083)
Rural Households (yes=1)	0.6389	0.7379	-0.0990
	(0.0040)	(0.0043)	(0.0082)
Scheduled Caste (yes=1)	0.1485	0.2078	-0.0594
	(0.0026)	(0.0043)	(0.0068)
Years of Education	5.7859	8.9379	-3.1521
	(0.0374)	(0.0450)	(0.0680)
Poorest Households (yes=1)	0.1061	0.2098	-0.1037
	(0.0021)	(0.0054)	(0.0068)

Notes: This table summarizes the results from the Classification Analysis. We present average characteristics of the individuals who were most and least affected by upwind fires. The two groups were defined based on the top and bottom 10% cutoffs shown in Figure 4. The last column shows the difference in averages between the two groups. The rows are sorted from highest to lowest percentage differences. Standard errors, in parentheses, were obtained with the bootstrap inference procedure from Chernozhukov, Fernández-Val, and Luo (2018). All differences are statistically significant at 1%, even after accounting for multiple hypotheses testing.

Appendix – For Online Publication

A Data Appendix

Data on hypertension incidence comes from the blood pressure tests for nearly 784,000 individuals from the National Family and Health Survey (NFHS) – IV (IIPS and ICF, 2017). The NFHS is a nationally representative survey covering nearly 600,000 households across India implemented between January 2015 to December 2016. It is the Indian wave of the globally implemented Demographic and Health Surveys (DHS). The NFHS-IV survey sample is representative at the district level across all 640 districts of the country. The survey interviewed all women aged 15 - 49 and for a subset of men aged 15 - 54 years within each sample household. Along with socio-economic, demographic, and fertility data, the NFHS also included biomarker data that included anthropometry measurements, anemia, blood glucose, blood pressure, and HIV testing. A vital advantage of the NFHS data for our research is the broad geographic coverage of the sample. The household-level blood pressure testing also implies that, unlike studies that often rely on hospital records or other health administrative data, our estimates are not biased selective reporting by only individuals whose health conditions were severe enough to seek medical attention. Such selection issues are particularly important in a developing country context such as India, where many individuals may not have access to medical care and, therefore, are not likely to show up in hospital data.

The NFHS measured blood pressure for all eligible men and women using trained field staff and standard medical equipment (an Omron Blood Pressure Monitor) to determine the prevalence of hypertension. For each individual, blood pressure readings were taken three times at five-minute intervals. We use standard cut-offs used in the medical literature to classify individuals with systolic blood pressure (BP)$\geq$ 140 mmHg or diastolic BP$\geq$ 90 mmHg as having hypertension (Prenissl et al., 2019). The NFHS also uses the same cut-off to identify individuals with elevated blood pressure readings and these respondents were encouraged to see a doctor for a full evaluation (IIPS and ICF, 2017).

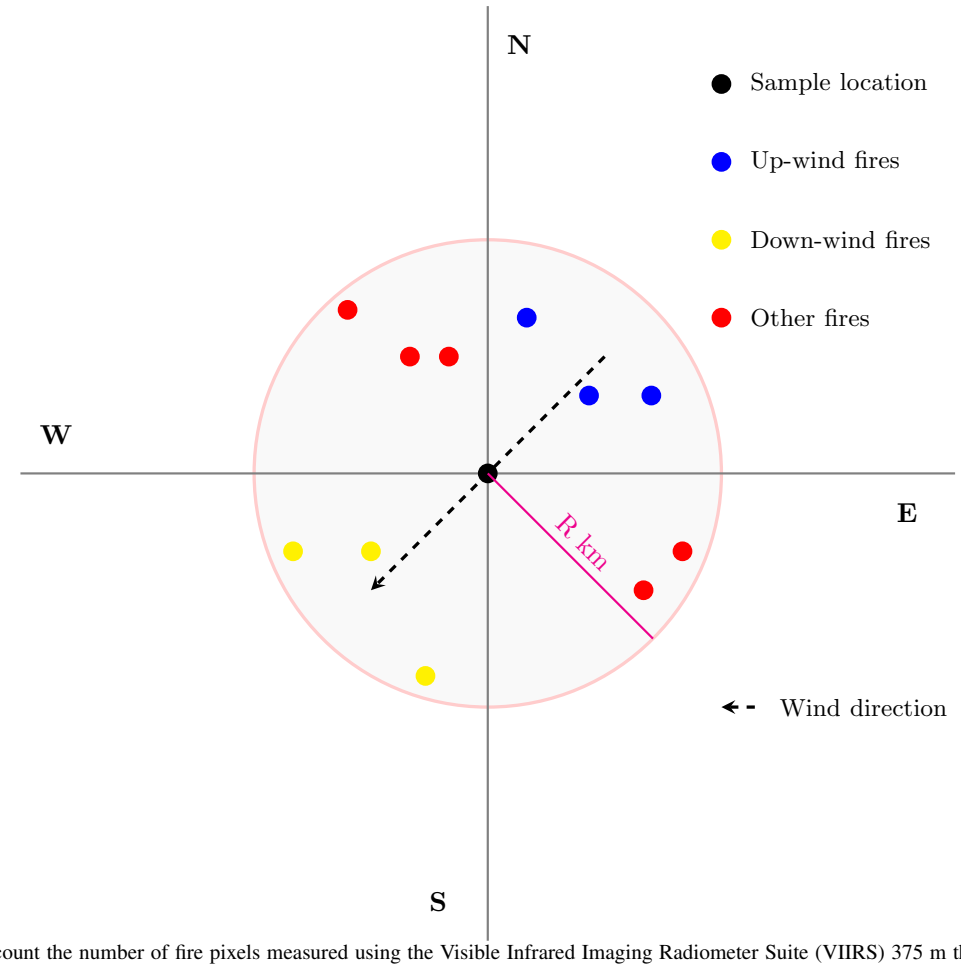
Along with the health test's date, the NFHS provides detailed information on individual and household characteristics, as well as geocoded information on the location of the household. The NFHS provides location information at the primary enumeration unit or sample cluster level, following the methodology followed by all Demographic and Health Survey globally. The population and size of sampled clusters vary by location. Typically, each cluster contains between 100 - 300 households, of which 20-30 households are randomly selected for survey participation. A cluster usually corresponds to a village in rural areas or a city block in urban areas. The cluster geo-coordinates data are assigned by first, taking the center of the sample enumeration area. Second, these codes are geomasked using a GPS coordinate displacement process to ensure the privacy of respondents (Burgert et al., 2013). Urban clusters are displaced up to 2-kilometers, and rural clusters by 10-kilometers.

We match the location and timing of the individual health measures to nearby fire activity at the daily level. We also match individual observations to daily location-specific pollution levels ($PM_{2.5}$, Black Carbon, Organic Carbon, SO_2, dust, and sea salt). In order to overcome the limitation posed by lack of ground monitoring data, we use satellite and model-derived pollution estimates from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2; Gelaro et al., 2017). Finally, we also match individuals to prevailing weather variables on temperature, precipitation, wind speed and direction using the ERA-5 climate reanalysis data (Hersbach et al., 2020). We aggregate hourly observations from the ERA-5 data to the daily level by calculating the average temperature and wind speed, and the total daily precipitation. Wind-direction is calculated as follows.

Figure A.1 below presents a schematic for our categorization of fires as upwind, downwind, or other directions. We aggregate the hourly ERA-5 data to a daily level by taking the vector-averaged daily mean wind direction at each respondent's location. Based on this daily mean, we classify the prevailing wind direction as falling in one of the 90-degree quadrants starting from 0-degree at the North. We create a 90-degree sector of the required radius around the respondent's location, corresponding to each wind quadrant. Matching these sector buffers with the daily fire-pixels data, we count the number of fires detected within each quadrant. In doing these calculations, we assume that the receptor pixel's wind direction (the respondent location) is the same as the source pixel (a fire location). We make this assumption to ease the computation involved in identifying up and downwind fires. While this is a simplifying assumption, the pollution results help validate these definitions of up and downwind fires: we find that upwind fires have a larger relative effect on pollution levels than downwind fires. We find that these patterns hold for pollution measured using modeled data (Figure 3 and Appendix D.1) and monitoring station data (Appendix D.2), and consistent with the pollution from fires being transported along the wind direction.

Table A.1 presets descriptive statistics for the main variables in this study: fires, ambient air pollution, and hypertension risk. Table A.2 presents descriptive statistics for the control variables considered in some regression specifications.

Figure A.1: Schematic illustrating definition of upwind fires



Notes: We count the number of fire pixels measured using the Visible Infrared Imaging Radiometer Suite (VIIRS) 375 m thermal anomalies and active fires data (EOSDIS, 2016) within a buffer of radius 50, 75, 100 or 150 km (shown by the shaded circle in the figure) around each individual's location in our sample. We then measure wind direction at the location using ERA-5 climate reanalysis data (Hersbach et al., 2020). The wind direction is then classified into 90-degree quadrants as shown above. Fire pixels that fall within the quadrant from which the wind is blowing towards the individual's location is classified as upwind fires. Fire pixels in the opposite quadrant constitute downwind fires, while remaining fires are classified as "other."

Table A.1: Descriptive Statistics for Main Variables of Interest

	Average	Standard Deviation	Min	Max
Number of Fires on Day Before Test				
Upwind fires within 50km	1.25	7.22	0.00	304.00
Downwind fires within 50km	1.11	5.85	0.00	305.00
Other fires within 50km	2.41	11.15	0.00	459.00
Upwind fires within 75km	2.83	13.89	0.00	518.00
Downwind fires within 75km	2.48	11.31	0.00	513.00
Other fires within 75km	5.27	21.26	0.00	593.00
Upwind fires within 100km	4.91	21.49	0.00	655.00
Downwind fires within 100km	4.37	17.42	0.00	651.00
Other fires within 100km	9.17	33.13	0.00	722.00
Upwind fires within 150km	9.82	35.77	0.00	849.00
Downwind fires within 150km	9.35	31.95	0.00	795.00
Other fires within 150km	18.93	55.80	0.00	845.00
Concentration of Pollutants on Day Before Test ($\mu g/m^3$)				
Particulate Matter ($PM_{2.5}$)	40.38	27.97	0.79	1796.91
Black Carbon	1.41	1.26	0.01	65.74
Organic Carbon	7.21	12.42	0.02	1069.50
SO2	5.94	5.53	0.00	87.18
Dust	16.28	14.75	0.01	194.34
Sea Salt	3.41	4.38	0.00	120.47
Blood Pressure (BP) Measures				
High BP ('000)	94.67	292.76	0.00	1000.00
– if not exposed to upwind fires	93.35	290.92	0.00	1000.00
Mildly Elevated or Worse BP	204.40	403.26	0.00	1000.00
– if not exposed to upwind fires	202.09	401.56	0.00	1000.00
Systolic BP (mmHg)	114.42	12.08	89.50	155.50
– if not exposed to upwind fires	114.41	12.02	89.50	155.50
Diastolic BP (mmHg)	76.53	8.80	57.00	104.00
– if not exposed to upwind fires	76.49	8.77	57.00	104.00
Number of Individuals	783,749			

Notes: This table presents averages, standard deviations, minimum, and maximum values for the main variables of interest for this study. We count number of fires that happened in the 24 hour period leading up to the day of the test. Fire counts are based on fire pixels observed using the VIIRS data (EOSDIS, 2016). Concentration of pollutants are based on estimates constructed from MERRA-2 climate reanalysis data (Gelaro et al., 2017). Blood pressure test data were obtained from the NFHS-IV survey (IIPS and ICF, 2017).

Table A.2: Descriptive Statistics for Control Variables

	Average	Standard Deviation	Min	Max
Wealth Index (categorical)	2.98	1.39	1.00	5.00
Age (years)	30.05	9.95	14.00	54.00
Education Level (yars completed)	6.98	5.12	0.00	20.00
Height (cm)	153.58	7.44	83.70	210.40
Weight (kg)	51.35	11.03	15.00	175.00
Arm Circumference (cm)	25.12	3.28	5.00	80.00
Married (%)	70.64	45.54	0.00	100.00
Smokes (%)	2.18	14.62	0.00	100.00
Prescribed BP Medicine (%)	2.92	16.84	0.00	100.00
Female (%)	86.29	34.39	0.00	100.00
Pregnant (%)	4.03	19.67	0.00	100.00
Household Size	5.62	2.61	0.00	40.00
Has Children (%)	39.05	48.79	0.00	100.00
Religion (categorical)	1.38	0.69	1.00	3.00
Scheduled Caste (%)	18.21	38.59	0.00	100.00
Schedulued Tribe (%)	17.80	38.25	0.00	100.00
Clean Cooking Fuel (%)	38.25	48.60	0.00	100.00
Outdoor Cooking (%)	68.79	46.33	0.00	100.00
Time to Get to Water Source (hours)	0.61	1.02	0.00	4.00
Has Ag. Land (%)	46.95	49.91	0.00	100.00
Rural Household (%)	70.73	45.50	0.00	100.00
Slept the Night Before BP Test (%)	98.32	12.87	0.00	100.00
Ate Before BP Test (%)	29.37	45.55	0.00	100.00
Had Coffee/Tea Before BP Test (%)	22.11	41.50	0.00	100.00
Smoked Before BP Test (%)	2.98	17.00	0.00	100.00
Other Tobacco Before BP Test (%)	6.53	24.70	0.00	100.00
Has Electric Fan (%)	73.31	44.23	0.00	100.00
Has AC (%)	19.82	39.86	0.00	100.00
Has Mosquito Net (%)	37.91	48.52	0.00	100.00
Health Insurance (%)	25.98	43.85	0.00	100.00
Hour of First BP Reading	13.72	2.80	2.00	23.00
Hour of Second BP Reading	13.78	2.80	2.00	23.00
Day of the Week (categorical)	3.01	2.00	0.00	6.00
Altitude of Location (m)	382.85	484.59	-5.00	5951.00
Rainfall on Test Day (mm)	3.80	9.53	0.00	261.19
Rainfall on Day Before Test (mm)	3.81	9.72	0.00	261.19
Temperature on Test Day (^oC)	27.28	6.30	-12.28	41.78
Temperature on Day Before Test (^oC)	27.23	6.35	-12.28	41.62
Wind Speed on Test Day (m/s)	2.79	1.78	0.01	14.77
Wind Speed on Day Before Test (m/s)	2.78	1.78	0.01	14.77
Wind Direction on Test Day (octants)	3.75	2.26	1.00	8.00
Wind Direction on Day Before Test (octants)	3.74	2.25	1.00	8.00
Number of Individuals	783,749			

Notes: This table presents averages, standard deviations, minimum, and maximum values for control variables included in the regression specifications. All variables were included as either binary or categorical in the regressions, except for age, education level, and the weather variables – rainfall,temperature, and wind speed and direction. To allow for non-linearities, variables that are typically continuous were transformed into categorical via binning. Rainfall, temperature, and wind speed are included as quadratic polynomials. The weather variables are drawn from ERA5 climate reanalysis data (Hersbach et al., 2020).

B Details on Empirical Strategy

B.1 Main Regression Specifications

As stated in the main text, we use reduced form regression specifications to estimate the impact of upwind agricultural fires on hypertension risk. The estimating equation can be written as:

$$\text{Reduced Form: } Y_{i,c,t} = \beta_1 Fires^{up}_{c,t-1} + \beta_2 Fires^{down}_{c,t-1} + \beta_3 Fires^{oth}_{c,t-1} + \boldsymbol{\psi}_x \mathbf{X}_{i,t} + \boldsymbol{\psi}_w \mathbf{W}_{c,t} + \boldsymbol{\theta}_{c,m} + \mu_{i,t}$$

where $Y_{i,c,t}$ is the blood pressure outcome for person i, located in NFHS cluster c and on the health survey day t; $Fires^{up}_{c,t-1}$ is the number upwind fires in the 24 hours leading to the blood pressure test within a specified buffer around cluster c; $Fires^{down}_{c,t-1}$ are downwind fires; $Fires^{oth}_{c,t-1}$ are fires from other wind directions.

We include district-by-month-of-sample fixed effects $\boldsymbol{\theta}_{c,m}$. These fixed effects flexibly account for district and season-specific factors that vary across space. Effectively, our specification compares blood pressure outcomes for individuals located within the same district and measured within the same month but exposed to different levels of pollution owing to variation in the number of upwind fires on the day of the blood pressure test.

In addition to the fixed effects, our regressions control for other factors that can affect pollution exposure or cardiovascular health. We control for local weather conditions $\mathbf{W}_{c,t}$ that affect the intensity of pollution exposure directly (e.g., if individuals choose to remain sheltered at home due to rain or extreme temperatures). We include wind direction, as well as quadratic polynomials of temperature, precipitation, wind speed on the day before and the day of the blood pressure test.

We also include a number of variables $\mathbf{X}_{i,t}$ that capture various socioeconomic characteristics such as wealth, education, marginalized caste or tribe groups that could be associated with pollution avoidance and health-seeking behavior. We also account for factors directly associated with individuals' health and blood pressure, such as age, weight, height, smoking behavior, and food consumed before blood pressure tests and other variables. Finally, we include indicators for day of the week in order to account for any changes pollution or behavior across specific days of the week (for instance, there might be changes in pollution due to lower traffic on weekends, individuals may be outside for longer due to work during weekdays or other factors). Note that the empirical strategy that we use does not require the inclusion of these control variables to estimate pollution's causal effect on hypertension and we get consistent results when we exclude these controls. Descriptive statistics for the full set of control variables are presented in Table A.2.

The equation is estimated via ordinary least squares (OLS). The parameter $\mu_{i,t}$ represents an idiosyncratic error terms. We cluster standard errors at the village (NFHS sample cluster) level to account for any within village heteroscedasticity. The main parameter of interest is β_1, relating hypertension risk with upwind fires.

Table C.1 shows the estimates from the above specification for using 50, 75, 100 and 150 km radii as buffers to measure fire exposure. Table C.3 shows the estimates from regressions where we maintain the fixed effects but exclude the additional control variables. Table C.4 presents results with alternative fixed effects. We find similar estimates across all these specifications. In Table C.5 we present estimates using an alternate measure of blood pressure risk - systolic blood pressure $\geq$ 130 mmHg or diastolic BP$\geq$ 85 mmHg. We find that exposure to upwind fires results in an elevated risk of high blood pressure using the alternate outcome measure as well. In Table C.6 we show estimates in which we use a 45-degree octant to define up and downwind instead of the 90-degree quadrant used in our main specification. Results remain consistent with our earlier estimates.

We also test if exposure on days prior or after the 24 hours leading to the test have any effect on hypertension. To do so, we estimate the reduced form model using fire counts exposure for the six days before and after the day of the blood pressure test. Figure C.1 presents these results. We see that the impact on hypertension is driven only by the short-term exposure in the 24 hours in the lead up to the test.

B.2 Sorted Effects Method and Classification Analysis

To assess heterogeneity of the effects of air pollution on cardiovascular distress, we implement the Sorted Effects Method and Classification Analysis developed by Chernozhukov, Fernández-Val, and Luo (2018). Both methods are based on the concept of partial effects. Suppose we observe a matrix with variables (Y, K, Z), where Y is a binary variable for incidence of high blood pressure, K is a continuous variable measuring number of upwind fires (our key covariate of interest), and Z are other controls. Lower case (y, k, z) represent realizations of those variables for a given observational unit. We can also define a binary response model $g(K, Z) = P[Y = 1|K, Z]$, representing the probability of high blood pressure, conditional on upwind fires and other factors.

In this setting, we are interested in the partial effects of changes in the number of upwind fires K, while holding other factors Z constant. These partial effects can be expressed as a partial derivative $\Delta(k, z) = \delta_k g(k, z)$, representing a marginal change in upwind fires from k holding other factors constant at z. Note that $\Delta(k, z)$ may be obtained for each observational unit. The first step for obtaining $\Delta(k, z)$ is to estimate a predictive model $\hat{g}(K, Z) = \hat{P}[Y = 1|K, Z]$. For that, we use the reduced form specification described in section B.1. One key difference is that, to allow for heterogeneous responses of the outcome to varying pollution levels, the reduced form is estimated via logistic regression, rather than OLS. Then it is possible to obtain predictions $\hat{P}[y = 1|k, z]$ for each observational unit.

The second step for estimating partial effects is to implement numeric differentiation with respect to k for the model that produces $\hat{P}[Y = 1|K, Z]$. We simulate marginal changes ε of k, such that the partial effects for each observational unit can be estimated as:

$$\hat{\Delta}(k, z) = \frac{\hat{P}[y = 1|k + \varepsilon, z] - \hat{P}[y = 1|k - \varepsilon, z]}{(k + \varepsilon) - (k - \varepsilon)} \quad .$$

We then rank the estimated partial effects $\hat{\Delta}(k, z)$ from lowest to highest. Finally, for estimation purposes and for validity of bootstrap inference, the tails from this ranking (top and bottom 5th percentiles) are dropped. Confidence intervals are produced via the bootstrap algorithm recommended by Chernozhukov, Fernández-Val, and Luo (2018). We run the estimation for 200 bootstrap iterations. Results for the Sorted Effects Method are discussed in the main text, and presented in Figure 4.

Classification Analysis consists of: first, identifying the subpopulations that are least and most affected by upwind fires, based on cutoffs for the percentiles of partial effects; second, estimating averages for covariates Z for the least and most affected subpopulations; third, taking the difference between the averages of each group. We select the bottom and top 10th percentiles as cutoffs. Standard errors for this analysis are also obtained via bootstrapping. Results are discussed in the main text, and presented in Table 1. All analyses from this section were implemented with the SortedEffects R package (Chen et al., 2020).

B.3 Attributable deaths and monetary costs calculation

Table B.1: Estimating mortality and mortality costs associated with high BP attributable to agricultural fire exposure

	(1)	(2)	(3)	(4)	(5)
		Source	Estimate	Upper bound	Lower bound
(A)	Total population (2015-16)	GBD	645,042,675	645,042,675	645,042,675
(B)	High BP prevalence	NFHS-IV	9.47%	9.53%	9.40%
(C)	Total number of high BP cases	$A \times B$	61,067,416	61,485,468	60,649,299
(D)	Deaths due to high BP (2015-16)	GBD	241,630	283,582	202,655
(E)	Share of high BP resulting in mortality	D/C	0.396%	0.461%	0.334%
(F)	Fraction of pop. exposed to ≥ 1 upwind fire	NFHS-IV and VIIRS	29.39%	29.49%	29.29%
(G)	Number of upwind fires on day before test	NFHS-IV and VIIRS	9.63	9.73	9.53
(H)	Marginal effect of upwind fires ('000)	Estimated (Table C.1)	0.121	0.205	0.037
(I)	High BP cases attributable to upwind fires (per exposure day)	$A \times F \times G \times H/1000$	220,817	379,784	66,093
(J)	Days in a year exposed to ≥ 1 upwind fire	NFHS-IV and VIIRS	37.80	37.87	37.72
(K)	High BP cases attributable to upwind fires (annual)	$I \times J$	8,346,168	14,382,907	2,493,178
(L)	Share of total high BP attributable to fires	K/C	13.67%	23.39%	4.11%
(M)	Additional mortality from High BP attributable to fires	$K \times E$	33,024	66,337	8,331
(N)	Annual costs of additional mortality (USD billion)	$M \times VSL$ \$	9.08	18.24	2.29

Notes: This table shows the procedure used to estimate the additional mortality from hypertension attributable to agricultural fires exposure and the associated monetary cost at the national level. Using estimates of the 2015-16 average population for India in the 20 - 54 age group from Global Burden of Disease Collaborative Network (2020a) and the percentage of individuals with high BP from the nationally representative NFHS-IV sample, we estimate the total number of high BP cases in the country in row (C). We then use the estimates of the total deaths in India attributable to high BP (average of 2015 and 2016) from the Global Burden of Disease (GBD) Model (Global Burden of Disease Collaborative Network, 2020b) to estimate the fraction of high BP cases that result in deaths (E). We assume that short-term changes in blood pressure are associated with as much risk of mortality as long-term changes in blood pressure (Dai et al., 2019). NFHS-IV data matched to the VIIRS fire measures provides us the share of the population exposed to at least 1 upwind fire in a 24-hour period (F), and the average number of upwind fires among exposed individuals during that period (G). We then multiply our main estimate from Table C.1 by the total population, the fraction of exposed individuals, and the number of upwind fires to which they were exposed to derive the daily country-level estimate of high BP cases attributable to fire exposure (I). We multiply this daily estimate by the average number of days in a year for which individuals are exposed (J), resulting in the annual estimate of high BP cases attributable to agricultural fire exposure (K). That allows us to obtain the share of high BP attributable to fires (L). Using the mortality risk from high BP from (E), we get the number of additional deaths arising due to hypertension from fire exposure (M). Finally, we use estimates of value of statistical life (VSL) for India from Viscusi and Masterman (2017) (0.275 million USD) to calculate the monetary costs associated with the additional hypertension-related mortality due to agricultural fire exposure (N).

C Further Results and Robustness

This section presents the main regression estimates of the impact of fires on hypertension, as well as the results of various robustness tests. Table C.1 shows the regression estimates from the main reduced form specification described in Appendix B.1. Each column in this table shows estimates from a separate OLS regression which corresponds to the results shown in Figure 2. In Table C.2, we show that results remain similar using a logit estimation instead of OLS. Panel A presents logit coefficients (log-odds), while Panel B presents "average parital effects" (APEs). For instance, the average partial effect of upwind fires for the 75 km specification in column (2) is 0.108, similar to the marginal effect of 0.121 obtained from the OLS model in column (2) of Table C.1. Our results also remain robust to the exclusion of weather or demographic controls in Table C.3, suggesting that the district-by-month-of-sample fixed effects may already capture most of the confounding variation in our observational data. We also see that the coefficient estimates remain unchanged when we use progressively more demanding sets of fixed effects in Table C.4.

Our results also hold when we change the definition of the outcome variable. In Table C.5, we change the outcome variable to be an indicator for "mildly elevated" BP, defined as systolic blood pressure $\geq$ 130 mmHg or diastolic BP$\geq$ 85 mmHg (compared to the primary outcome variable, which was an indicator for systolic blood pressure $\geq$ 140 mmHg or diastolic BP$\geq$ 90 mmHg). The marginal effect of upwind fires remains similar to the estimate obtained in the main specification. Finally, we find that the results are not sensitive to the choice of how "upwind" is defined. In Table C.6, we classify upwind and downwind fires based on 45-degree octants (rather than 90-degree quadrants used in the main specification). Point estimates for upwind fires are slightly larger in magnitude compared to the main specification, but the effect size range remains qualitatively similar.

In sub-section C.4 we tested for lead and lagged effects of upwind fires on hypertension. To do so, we augment the main model specification described in B.1 with additional terms for fire counts in up, down, and other directions for $\pm$6 days. We also include weather controls for all $\pm$6 days. The remaining individual, demographic, and household covariates included in the main model remain unchanged. Figure C.1 presents the coefficients on upwind fires within 75-kilometer corresponding to the days leading up to and after the blood pressure test (Table C.7 shows the corresponding regression estimates). We see that upwind fires from more than one day before the tests or days after the tests do not significantly affect hypertension risk.

Sub-section C.5 presents results using a continuous outcome measure instead of the binary outcome used in the main results. We re-estimate our main regression specification with the outcome changed to systolic blood pressure in Panel (a) and diastolic blood pressure in Panel (b), respectively, of Figure C.2. The results shown are for the 75 km exposure measure. For each outcome, we estimate the regression for the full sample (left-most plot in each panel) and separately for low risk (middle plot) and high risk (right-most plot) sub-groups in the sample. We see that the coefficients on downwind fires are null across all specifications. Conversely, we find positive and statistically significant effects for upwind fires, but only for subsamples of "high risk" individuals who had systolic BP $\geq$ 120; diastolic BP $\geq$ 80. Additionally, heterogeneity analyses in Figure C.3 suggest stronger effects for individuals with body mass index (BMI) $\geq$ 30, or with age $\geq$ 40. The coefficients in Figure C.3 are obtained from the interactions of upwind fires with given categorical variables of interest.

Finally, sub-section C.6 presents descriptive evidence in support of the primary identifying assumption: daily wind direction is quasi-random, leading to a plausibly exogenous variation in exposure to the number of upwind agricultural fires. Figure C.4 plots, for each interview date, the shares of NFHS clusters that experienced wind direction changes compared to the day prior to the interview. Panel (a) measures changes based on wind direction quadrants, while Panel (b) measures changes based on octants. The plots show that there is substantial day-to-day variation in wind directions during the NFHS survey period. In about 33% of the cases, wind direction quadrants on the interview day were different from those on the day before. This share is even higher (48%) if we define wind direction based on octants. The implication is that farmers are unlikely to be able to predict day-to-day changes in wind direction and thus have little scope to manipulate our primary exposure variable (upwind fires).

C.1 Robustness of results to alternative regression specifications

Table C.1: Impact of exposure to agricultural fires (day before test) within varying distances around location on the incidence of hypertension

	Hypertension (per '000)			
Radius	50 km	75 km	100 km	150 km
	(1)	(2)	(3)	(4)
Fire counts				
Up-wind	0.137*	0.121***	0.069**	0.029*
	(0.083)	(0.043)	(0.028)	(0.016)
Down-wind	0.094	0.027	-0.002	-0.019
	(0.081)	(0.047)	(0.030)	(0.016)
Other directions	-0.102*	-0.070**	-0.033*	-0.008
	(0.052)	(0.029)	(0.018)	(0.010)
'Control' average	93.810	93.350	93.010	93.320
R^2	0.11	0.11	0.11	0.11
Observations	783,773	783,773	783,773	783,773
Fixed-effects				
District × Month of sample	✓	✓	✓	✓
Day of week	✓	✓	✓	✓

Notes: Table shows estimates from separate OLS regressions in each column using 50, 75, 100, and 150 km radius buffers for fire counts. The dependent variable is a binary variable for incidence of hypertension taking the value 1000 if systolic blood pressure $\geq$ 140 mmHg or diastolic BP$\geq$ 90 mmHg and zero otherwise. The row 'Control' average shows the mean of the outcome variable for the sample of individuals that were not exposed to any upwind fires within each buffer radius. Up and downwind is based on 90-degree wind sectors. All specifications include district-by-month of sample and day of week fixed effects. Specifications further control for weather variables (temperature, rainfall, wind-speed and wind-direction), household and individual characteristics. Standard errors shown in parentheses, clustered at NFHS sample cluster level Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table C.2: Impact of exposure to agricultural fires (day before test) – Logit regression estimates

Radius	50 km	75 km	100 km	150 km
	(1)	(2)	(3)	(4)
	Panel A: Likelihood of Hypertension			
Fire counts				
Up-wind	0.0015	0.0015***	0.0008**	0.0003*
	(0.0010)	(0.0005)	(0.0003)	(0.0002)
Down-wind	0.0013	0.0005	0.0000	-0.0002
	(0.0010)	(0.0006)	(0.0004)	(0.0002)
Other directions	-0.0013**	-0.0009***	-0.0004*	-0.0001
	(0.0006)	(0.0004)	(0.0002)	(0.0001)
'Control' average	93.8100	93.3500	93.0100	93.3200
Pseudo R^2	0.16	0.16	0.16	0.16
Observations	783,144	783,144	783,144	783,144
Fixed-effects				
District × Month of sample	✓	✓	✓	✓
Day of week	✓	✓	✓	✓
	Panel B: Average Partial Effects (per '000)			
Fire counts				
Up-wind	0.111*	0.108***	0.062***	0.025**
	(0.057)	(0.034)	(0.020)	(0.011)
Down-wind	0.100	0.037	0.002	-0.021
	(0.071)	(0.046)	(0.026)	(0.013)
Other directions	-0.099***	-0.073***	-0.036*	-0.011
	(0.037)	(0.026)	(0.019)	(0.009)
'Control' average	93.810	93.350	93.010	93.320

Notes: Table shows estimates from separate logit regressions in each column using 50, 75, 100, and 150 km radius buffers for fire counts. The dependent variable is a binary variable for incidence of hypertension taking the value 1 if systolic blood pressure $\geq$ 140 mmHg or diastolic BP$\geq$ 90 mmHg and zero otherwise. Estimates shown in Panel A are logit coefficients (log-odds). Taking the exponent of these coefficients yields the "odds-ratio". For example, in column (2), a one s.d. (13.9) increase in upwind fires within 75 km increases the odds of hypertension by 1.02 ($\approx e^{0.0015\times 13.9}$). Panel B shows "average partial effects" (APEs) estimated with the procedure from Chernozhukov, Fernández-Val, and Luo (2018). These APEs represent effects on incidence of hypertension per thousand ('000), and are comparable to the OLS estimates from Table C.1. 'Control' average shows the mean of the outcome variable for the sample of individuals that were not exposed to any upwind fires within each buffer radius. Up and downwind is based on 90-degree wind sectors. All specifications include district-by-month of sample and day of week fixed effects. Specifications further control for weather variables (temperature, rainfall, wind-speed and wind-direction), household and individual characteristics. Standard errors, shown in parentheses, were clustered at NFHS sample cluster level for Panel A. Standard errors for Panel B were bootstrapped. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table C.3: Impact of exposure to agricultural fires (day before test) on the incidence of hypertension – Robustness to excluding control variables

	Hypertension (per '000)			
Radius	50 km	75 km	100 km	150 km
	(1)	(2)	(3)	(4)
Fire counts				
Up-wind	0.150*	0.134***	0.085***	0.035**
	(0.080)	(0.045)	(0.029)	(0.016)
Down-wind	0.040	-0.014	-0.022	-0.030*
	(0.084)	(0.049)	(0.031)	(0.017)
Other directions	-0.109**	-0.073**	-0.045**	-0.018
	(0.054)	(0.031)	(0.019)	(0.011)
'Control' average	93.810	93.350	93.010	93.320
R^2	0.02	0.02	0.02	0.02
Observations	783,773	783,773	783,773	783,773
Fixed-effects				
District × Month of sample	✓	✓	✓	✓
Day of week	✓	✓	✓	✓

Notes: Table shows estimates from separate OLS regressions in each column using 50, 75, 100, and 150 km radius buffers for fire counts. The dependent variable is a binary variable for incidence of hypertension taking the value 1000 if systolic blood pressure $\geq$ 140 mmHg or diastolic BP$\geq$ 90 mmHg and zero otherwise. The row 'Control' average shows the mean of the outcome variable for the sample of individuals that were not exposed to any upwind fires within each buffer radius. Up and downwind is based on 90-degree wind sectors. All specifications include district-by-month of sample and day of week fixed effects, but exclude any other control variables. Standard errors shown in parentheses, clustered at NFHS sample cluster level Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table C.4: Impact of exposure to agricultural fires within 75-kilometers (day before test) on the incidence of hypertension – Robustness to alternative fixed effects

	Hypertension (per '000)			
	(1)	(2)	(3)	(4)
Fire counts				
Up-wind	0.121***	0.136***	0.109**	0.109*
	(0.043)	(0.049)	(0.054)	(0.063)
Down-wind	0.027	0.021	0.055	0.043
	(0.047)	(0.050)	(0.061)	(0.072)
Other directions	-0.070**	-0.095***	-0.075*	-0.066
	(0.029)	(0.032)	(0.040)	(0.044)
R^2	0.105	0.109	0.153	0.163
Observations	783,775	783,775	783,775	783,775
Fixed-effects				
District × Month of sample	✓			
Day of week	✓	✓	✓	✓
District × Week of sample		✓		
NFHS cluster × Month of sample			✓	
NFHS cluster× Week of sample				✓

Notes: Table shows estimates from separate OLS regressions in each column. The dependent variable is a binary variable takes value 1, 000 if individual has hypertension (systolic blood pressure $\geq$ 140 mmHg or diastolic BP$\geq$ 90 mmHg) and zero otherwise. Up and downwind fires based on 90-degree wind sectors, counting fire-pixels within 75-kilometers of individuals location. Fixed effects used vary by specification across columns. Column (1) repeats the results of the main specification from Table C.1. All specifications include controls for weather (temperature, rainfall, wind-speed and wind-direction), household and individual characteristics. Standard errors shown in parentheses, clustered at NFHS sample cluster level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

C.2 Robustness of results to alternative definition of hypertension

Table C.5: Impact of exposure to agricultural fires (day before test) on the incidence of mildly elevated BP or worse

Panel A: Main specification			
Radius	75 km	100 km	150 km
	(1)	(2)	(3)
Fire counts			
Up-wind	0.132**	0.049	0.027
	(0.056)	(0.037)	(0.021)
Down-wind	0.071	0.008	-0.025
	(0.074)	(0.047)	(0.026)
Other directions	-0.044	0.007	0.016
	(0.045)	(0.029)	(0.017)
'Control' average	202.090	201.430	202.190
R^2	0.15	0.15	0.15
Observations	783,773	783,773	783,773
Fixed-effects			
District × Month of sample	✓	✓	✓
Day of week	✓	✓	✓
Panel B: Excluding weather, household and individual controls			
Radius	75 km	100 km	150 km
	(1)	(2)	(3)
Fire counts			
Up-wind	0.132**	0.049	0.027
	(0.056)	(0.037)	(0.021)
Down-wind	0.071	0.008	-0.025
	(0.074)	(0.047)	(0.026)
Other directions	-0.044	0.007	0.016
	(0.045)	(0.029)	(0.017)
'Control' average	202.090	201.430	202.190
R^2	0.15	0.15	0.15
Observations	783,773	783,773	783,773
Fixed-effects			
District × Month of sample	✓	✓	✓
Day of week	✓	✓	✓

Notes: Table shows estimates from separate OLS regressions in each column using 75, 100, and 150 km radius buffers for fire counts. The dependent variable is incidence of mildly elevated BP or worse (systolic blood pressure $\geq$ 130 mmHg or diastolic BP$\geq$ 85 mmHg) per $'000$. The third to last row shows the 'control' average outcome for the sample of individuals that were not exposed to upwind fires. Up and downwind is based on 90-degree wind sectors. All specifications include district-by-month of sample and day of week fixed effects. Specifications in Panel A further control for weather variables (temperature, rainfall, wind-speed and wind-direction), household and individual characteristics. Standard errors shown in parentheses, clustered at NFHS sample cluster level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

C.3 Robustness of results to alternative definition of wind direction

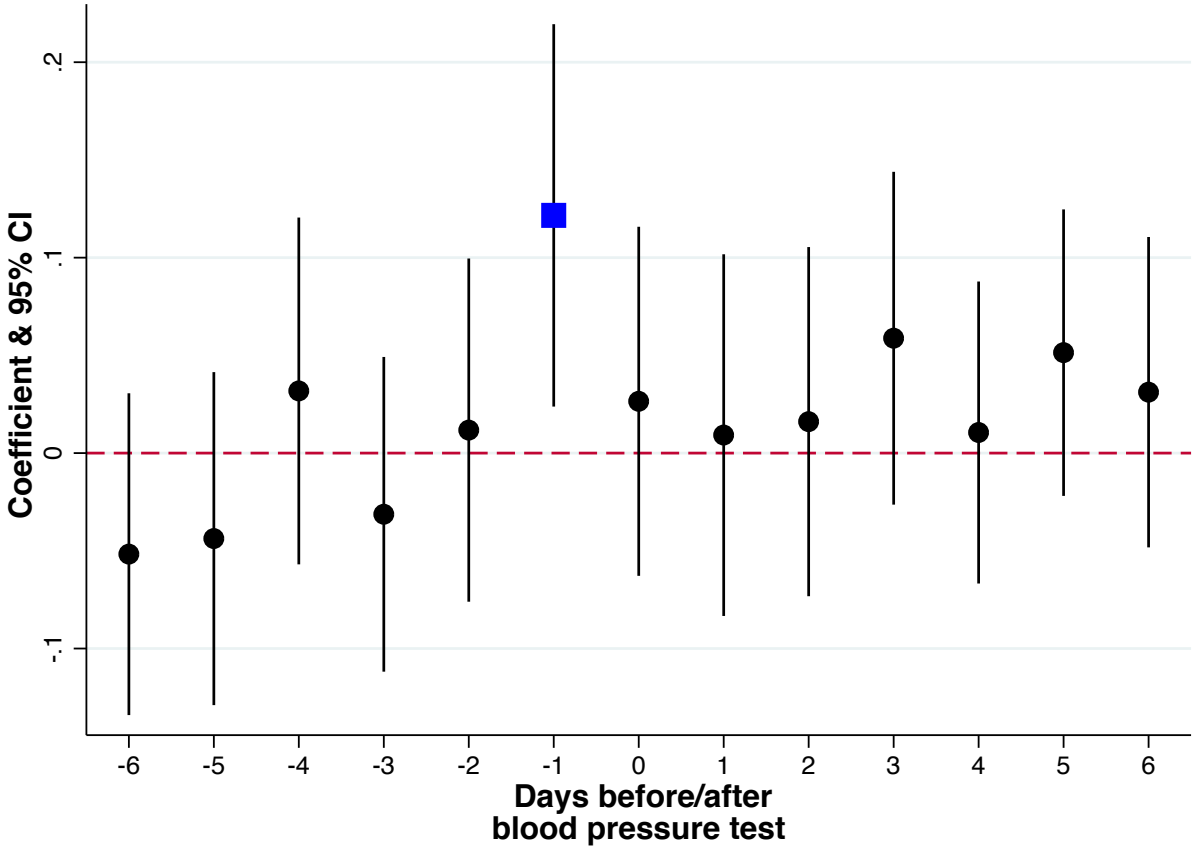
Table C.6: Impact of exposure to agricultural fires (day before test) on the incidence of hypertension – Robustness to change in wind-sector used to define up- and downwind.

	Hypertension (per '000)			
Radius	50 km	75 km	100 km	150 km
	(1)	(2)	(3)	(4)
Up-wind Fires	0.271*	0.185**	0.119**	0.027
	(0.151)	(0.079)	(0.050)	(0.028)
Down-wind fires	0.166	0.050	-0.042	-0.062**
	(0.141)	(0.080)	(0.052)	(0.029)
Other directions	-0.058	-0.030	-0.010	0.004
	(0.035)	(0.020)	(0.013)	(0.008)
'Control' average	94.14	93.82	93.73	93.73
Observations	783,773	783,773	783,773	783,773
R-squared	.1053	.1053	.1053	.1053
Fixed-effects				
District × Month of sample	✓	✓	✓	✓
Day of week	✓	✓	✓	✓

Notes: Table shows estimates from separate OLS regressions in each column using 50, 75, 100, and 150 km radius buffers for fire counts. The dependent variable is incidence of hypertension (systolic blood pressure $\geq$ 140 mmHg or diastolic BP$\geq$ 90 mmHg) per $'000$. The third to last row shows the 'control' average hypertension for the sample of individuals that were not exposed to upwind fires. Up and downwind is based on 45-degree wind sectors. All specifications include district-by-month of sample and day of week fixed effects, and further control for weather variables (temperature, rainfall, wind-speed and wind-direction), household and individual characteristics. Standard errors shown in parentheses, clustered at NFHS sample cluster level.. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

C.4 Lead and lagged effects

Figure C.1: Impact of upwind fire exposure on days before and after the day of blood pressure test on incidence of hypertension



Notes: The graph plots the coefficient on upwind fires within 75-kilometer on days leading up to and after the blood pressure test. Estimates shown are from a single regression. We augment the main model specification described in B.1 with additional terms for fire counts in up, down and other directions for ± 6 days. Regression includes controls for weather variables for all days and other covariates included in the main model. The vertical lines represent 95% confidence intervals that account for clustering at the village (NFHS cluster) level. We see that the impact on hypertension is primarily driven by the exposure during the 24 hour period prior to the test (the coefficient on -1 in the graph). We cannot the reject the null of zero for the coefficients on other days.

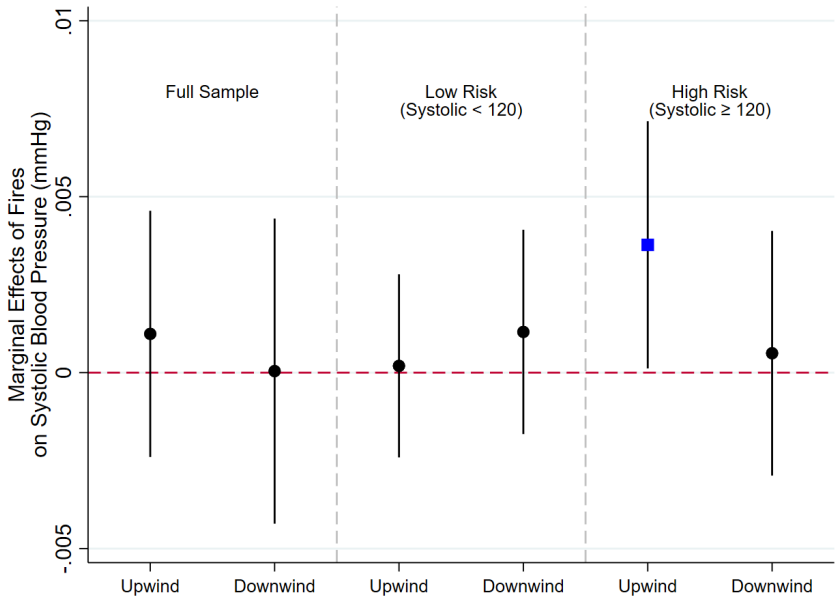
Table C.7: Impact of upwind fire exposure on days before and after the day of blood pressure test on incidence of hypertension

	Effect of days before(-)/after(+)	
	(1)	
	Coeff.	SE
Days from test:		
-6	-0.052	(0.042)
-5	-0.044	(0.043)
-4	0.032	(0.045)
-3	-0.031	(0.041)
-2	0.012	(0.045)
-1	0.122**	(0.050)
0	0.026	(0.046)
1	0.009	(0.047)
2	0.016	(0.046)
3	0.059	(0.043)
4	0.011	(0.039)
5	0.051	(0.037)
6	0.031	(0.041)
Observations	783,773	
R-squared	.1056	

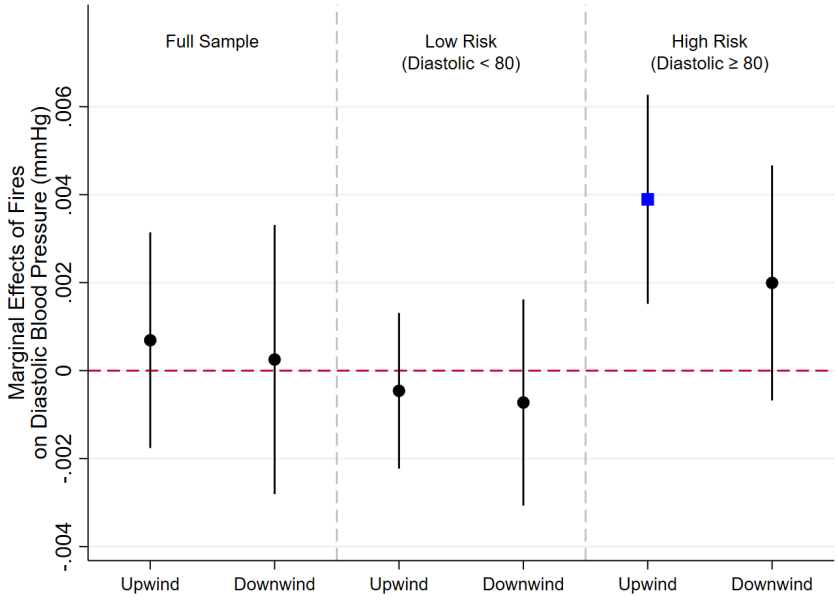
Notes: Table shows estimates the coefficient (in column 1) on upwind fires within 75-kilometer on days leading up to and after the blood pressure test. Estimates shown are from a single regression. We augment the main model specification described in B.1 with additional terms for fire counts in up, down and other directions for ± 6 days. Regression includes controls for weather variables for all days and other covariates included in the main model. Standard errors clustered at NFHS clustered are shown in parentheses in column (2). Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

C.5 Results with continuous measures of blood pressure

Figure C.2: Impact of exposure to fires on continuous measures of blood pressure



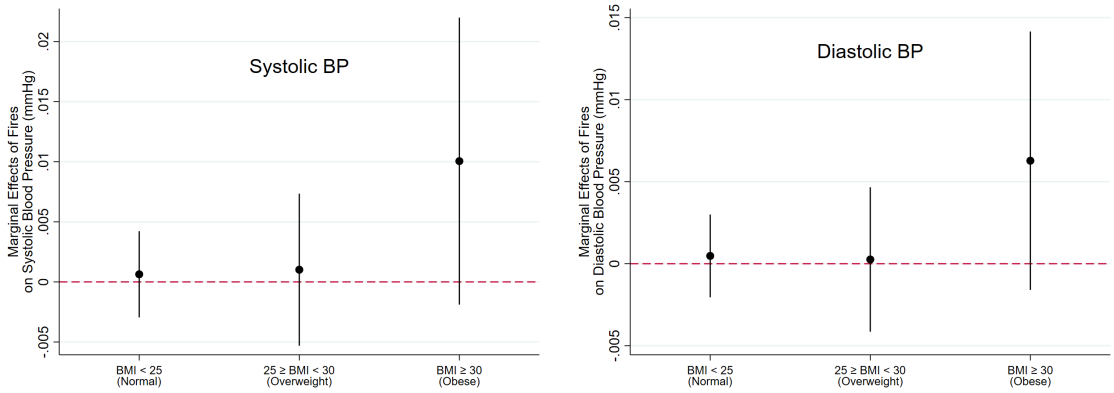
(a) Systolic Blood Pressure



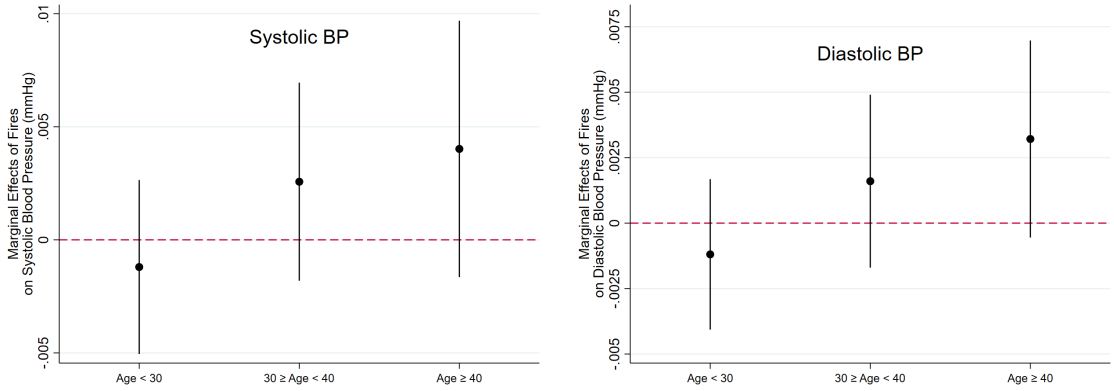
(b) Diastolic Blood Pressure

Notes: Plots show the marginal effects and 95% CIs of up- and downwind fires on continuous measures of blood pressure. We run separate regressions for Systolic BP (Panel a), and for Diastolic BP (Panel b), both measured in $mmHg$. All specifications use 75 km radius buffers for fire counts, include district-by-month of sample and day of week fixed effects, weather controls, as well as demographic and household characteristics. Up and downwind is based on 90-degree wind sectors. Standard errors are clustered at the NFHS sample cluster level. The two leftmost coefficients are for regressions with the full survey sample. The two coefficients in the center are from regressions restricted to individuals with normal BP levels (i.e., low risk). The two rightmost coefficients are for individuals with abnormal BP levels (i.e., high risk).

Figure C.3: Impact of exposure to fires on continuous measures of blood pressure – heterogeneity by risk factors



(a) Heterogeneity by BMI

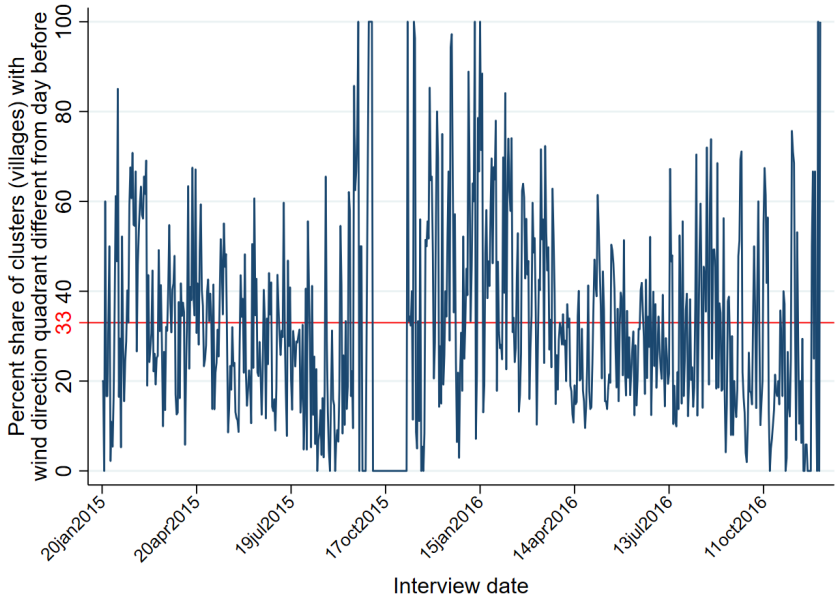


(b) Heterogeneity by Age

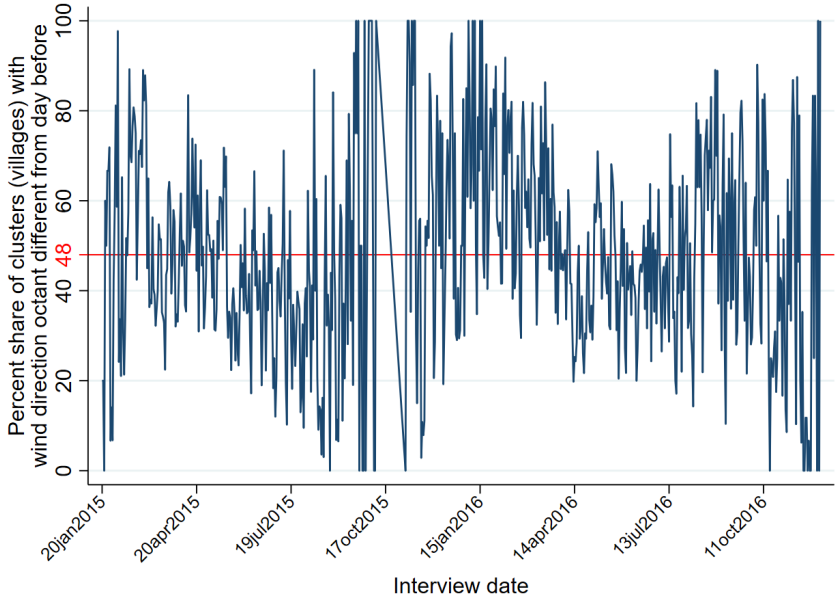
Notes: Plots show the marginal effects and 95% CIs of upwind fires on continuous measures of blood pressure. We run separate regressions for Systolic BP, and for Diastolic BP, both measured in $mmHg$. Panel (a) shows coefficients from the upwind fire counts interacted with indicators for BMI categories. Panel (b) is for interactions with Age categories. All specifications use 75 km radius buffers for fire counts, include district-by-month of sample and day of week fixed effects, weather controls, as well as demographic and household characteristics. Upwind is based on 90-degree wind sectors. Standard errors are clustered at the NFHS sample cluster level.

C.6 Quasi-randomness of wind directions

Figure C.4: Shares of clusters where wind direction from day of interview was different from that of the day before



(a) Changes in Quadrants of Wind Direction



(b) Changes in Octants of Wind Direction

Notes: These figures plot, for each interview date, the shares of NFHS clusters that experienced wind direction changes compared to the day prior to the interview. Panel (a) measures changes based on wind direction quadrants, while Panel (b) measures changes based on octants. The red horizontal lines represent the average shares of day-to-day wind direction changes across all clusters and throughout the whole survey period.

D Impact of Fires on Air Pollution

We hypothesize that fire-induced air pollution is the key mechanism linking exposure to fires to increased risk of hypertension. In this section, we present evidence in support of this mechanism. In sub-section D.1, we examine the impact of fires on model-based air pollutants using pollutant measures derived from reanalysis data. These outcome measures are satellite and model-derived pollution estimates from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (Gelaro et al., 2017:MERRA-2,). We aggregate the hourly observations to 24-hourly averages and match these pollution measures to each location-test date in our health data sample. We obtain surface concentrations of sulfate, organic carbon, black carbon, dust, and sea-salt particulate matter with a diameter of less than 2.5 μm. We combine these species level estimates to obtain total $PM_{2.5}$ concentrations using ground-validated conversion factors from He et al. (2019).

We use the same estimating equation as that used for analyzing the effect of fires on hypertension (described in Appendix B.1, excluding the individual-level covariates. Table D.1 shows these regression estimates. We see that upwind fires have a disproportionately larger impact on concentrations of $PM_{2.5}$, black carbon, organic carbon, and SO_2 (Columns (1) - (4)). Dust and sea-salt particulate concentrations are determined by long-range atmospheric transport and proximity to arid or desert regions and are less likely to be affected by fires. Consequently, we see that the impacts of fires on dust and sea salt particulate matter are close to zero (Columns (5) and (6)). The estimates shown in Figure 3 of the main text correspond to the regression estimates in Table D.1. The coefficients in Figure 3 are standardized to show the percentage change relative to the sample mean of the outcome variable to allow for easier comparison of the relative magnitude of fires impact across each of the pollutant outcomes.

These results rely on modeled pollution estimates that may be prone to measurement error. Reanalysis data such as MERRA-2 also often rely on assumptions about emission conversion factors and approximations in the underlying atmospheric chemical transport models, leading to biased predictions in locations with poor ground validation data. Therefore, to test if our pollution results are robust to such mismeasured outcomes, we also examine the impact of fires on pollution using daily pollution data from ground monitors in sub-section D.2.

We use daily monitor level data corresponding to the analysis period 2015-16 publicly available from the Central Pollution Control Board, Government of India (https://cpcb.nic.in/). The network of pollution monitoring stations is sparse in India and limited to urban areas. For the 2015-16 period, data were available for 83 monitoring stations spread across ten cities. We aggregate station data to the city-day level. Figure D.1 shows the location of these cities. We match cities to daily counts of fires in the up, down, and other directions. We also obtain daily weather variables to use as controls. Table D.2 presents regression estimates from a specification that includes city, day of the sample, and city-by-month fixed effects, and further control for weather variables (temperature, rainfall, wind speed, and wind direction). We see that upwind fires significantly increase levels of $PM_{2.5}$, PM_{10}, NO, and NO_2 (Columns (1) - (4)), with a positive, but imprecise effect on ozone (Column (5)). Downwind fires do not increase any of the pollutants. Figure D.2 presents the marginal effects as a percent change relative to the outcome sample mean on zero upwind fire days. We find the magnitude of impact on $PM_{2.5}$ remains similar to that obtained using modeled $PM_{2.5}$ data in Figure 3.

D.1 Pollution measures from satellite data

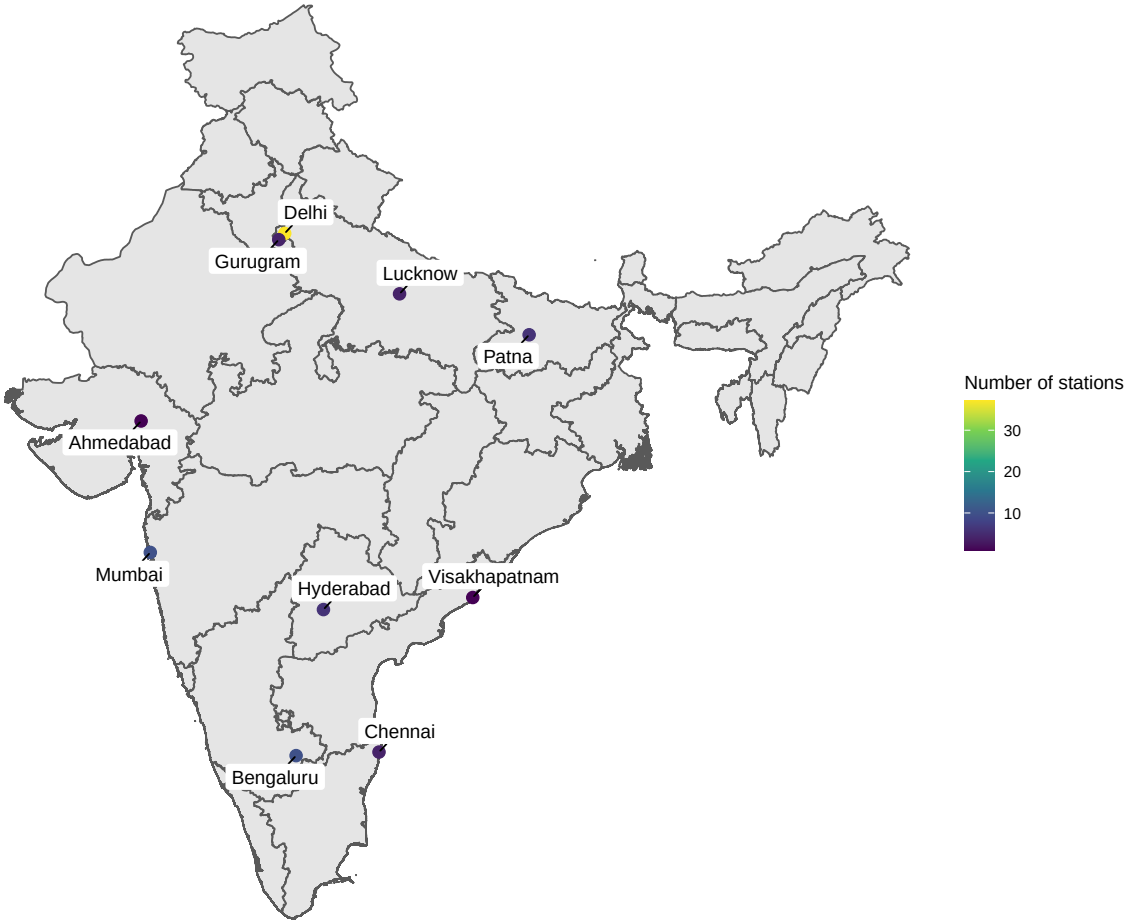
Table D.1: Impact of agricultural fires on daily air pollution measured using satellite data

	$PM_{2.5}$ (1)	BC (2)	OC (3)	SO_2 (4)	Dust (5)	Sea Salt (6)
Fire counts						
Up-wind	0.748***	0.032***	0.453***	0.039***	-0.017**	-0.003***
	(0.238)	(0.009)	(0.143)	(0.011)	(0.008)	(0.001)
Down-wind	0.263	0.013	0.178*	0.025***	-0.016	-0.000
	(0.173)	(0.008)	(0.102)	(0.009)	(0.011)	(0.002)
'Control' average	38.564	1.335	6.274	5.714	16.202	3.571
R^2	0.53	0.70	0.54	0.77	0.59	0.71
Observations	783,773	783,773	783,773	783,773	783,773	783,773
Fixed-effects						
Day of week	✓	✓	✓	✓	✓	✓
District × Month of sample	✓	✓	✓	✓	✓	✓

Notes: Table shows estimates from separate OLS regressions in each column using. The dependent variable are daily average pollution levels (in $\mu g/m^3$) at NFHS cluster locations obtained by aggregating hourly observations from MERRA-2 atmospheric reanalysis data. Up and downwind daily fires are counts of fire-pixels within 75 km around each location. MERRA-2 does not directly estimate $PM_{2.5}$. Instead, we obtain $PM_{2.5}$ by aggregating BC (Black Carbon), OC (Organic Carbon), SO_2 (Sulfate), Dust and Sea salt particles that together constitute $PM_{2.5}$ by using conversion factors from He et al. (2019). All regressions control for weather variables (temperature, rainfall, wind-speed and wind-direction). 'Control' average is the mean of the outcome variable for the sample with zero upwind fires. Standard errors clustered at the NFHS sample cluster level are shown in parentheses. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

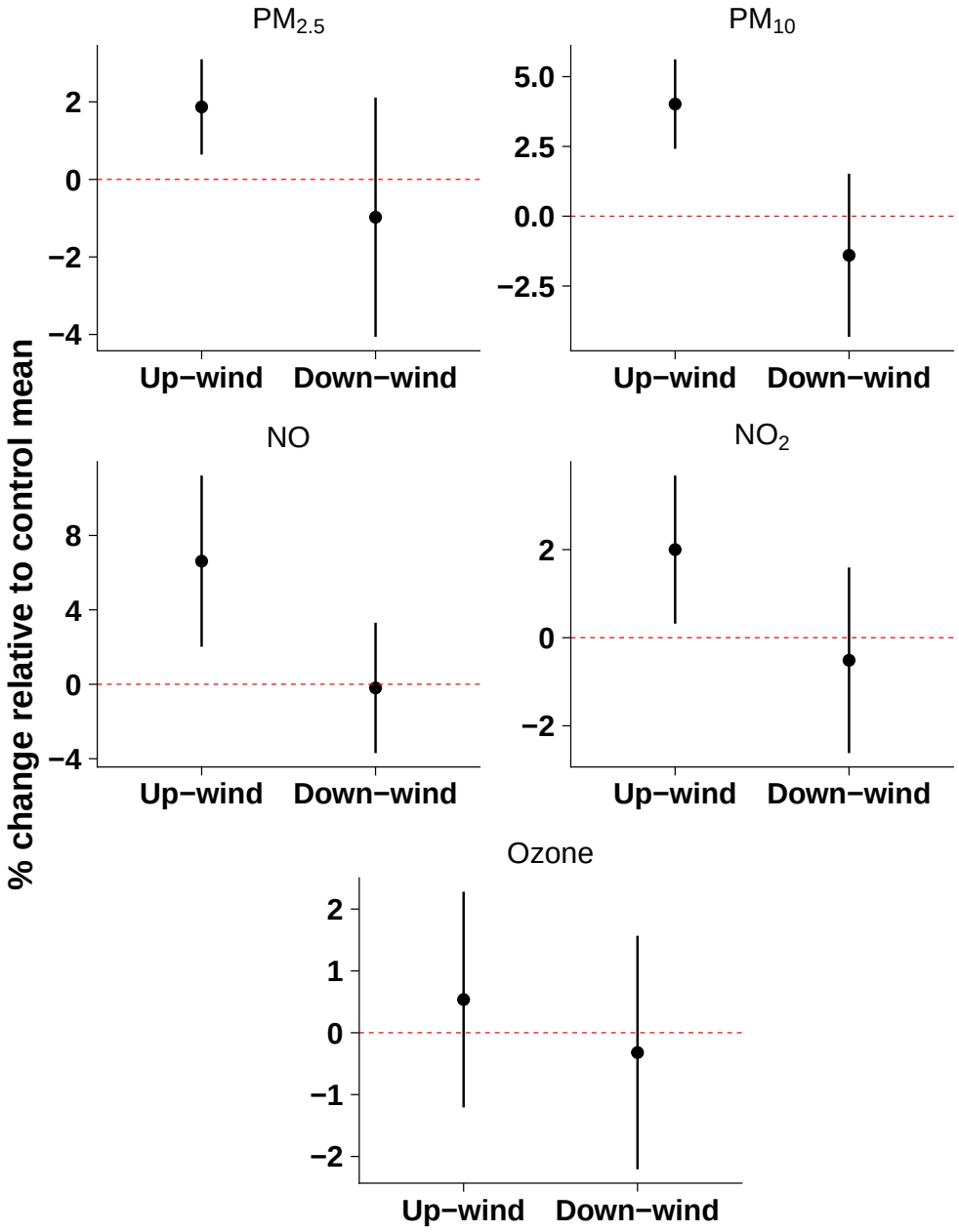
D.2 Pollution measures from ground monitors

Figure D.1: Location of air pollution monitoring stations



Notes: Map shows the location of the cities with data available for air pollution monitors for 2015-16 period from the Central Pollution Control Board, Government of India. A total of 83 stations were available across 10 cities. We aggregate hourly station-level data to the average city-day level. Given the limited availability of monitors, we cannot use these data in our main estimates across the NFHS sample.

Figure D.2: Effects of agricultural fires on air pollution measured at ground monitors (relative to "control" group mean)



Notes: Plots show the marginal effects of up- and downwind agricultural fires within 75-kilometers around ground-level air pollution monitors. The sample consists of daily pollutant measures for 10 cities shown in Figure D.1. The dependent variables are daily city-level average measure of the pollutant, obtained by aggregating hourly observations from all monitoring stations within each city (in $\mu g/m^3$). Coefficients are transformed to show the percentage change relative to the mean of the outcome variable on days with zero upwind fires ("control" group mean). Error bars show the associated 95% confidence intervals clustered at the city $\times$ day level. Each figure shows estimates from a separate regression. Up and downwind fires are counts of fire-pixels within 75 km around each city based on 90-degree wind sectors, excluding fire-pixels detected within 10 km of the city geocoordinates. All regressions control for rainfall, temperature, wind speed, and include city-by-month, day and city fixed effects.

Table D.2: Impact of agricultural fires on daily air pollution measured at ground monitors

	$PM_{2.5}$ (1)	PM_{10} (2)	NO (3)	NO_2 (4)	O_3 (5)
Fire counts					
Up-wind	1.565***	5.662***	1.112***	0.526**	0.181
	(0.525)	(1.152)	(0.395)	(0.225)	(0.301)
Down-wind	-0.818	-1.973	-0.033	-0.135	-0.108
	(1.319)	(2.103)	(0.300)	(0.282)	(0.326)
'Control' average	83.754	140.955	16.807	26.276	33.807
R^2	0.71	0.91	0.60	0.68	0.61
Observations	4,500	1,669	5,029	5,034	4,955

Notes: Table shows estimates from separate OLS regressions in each column using. The sample consists of daily pollutant measures for 10 cities shown in Figure D.1. The dependent variables are daily city-level average measure of the pollutant, obtained by aggregating hourly observations from all monitoring stations within each city (in $\mu g/m^3$). Not all pollutant are measured every day, and therefore the number of observations varies across outcomes. Up and downwind fires are counts of fire-pixels within 75 km around each city based on 90-degree wind sectors, excluding fire-pixels detected within 10 km of the city geocoordinates. All specifications include city, day of sample, and city-by-month fixed effects, and further control for weather variables (temperature, rainfall, wind-speed and wind-direction). 'Control' average is the mean of the outcome variable for the sample with zero upwind fires. Standard errors clustered at the city-day level are shown in parentheses. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

E Instrumental Variable Approach

As shown in Figure 3, we find that that upwind fires significantly increase the concentrations of $PM_{2.5}$, and its subcomponents, within 75km surrounding NFHS respondents' locations. This significant association, along with the results from section 2.1, may lead one to consider using upwind fires as an instrumental variable (IV) to estimate the direct effect of $PM_{2.5}$ exposure on hypertension risk. However, as discussed in section 2.2, IV estimates in this setting must be interpreted with caution. We argue that these estimates might not be valid due to violations of the exclusion restriction: i.e., fire activity could lead to increased concentration of other pollutants, not only $PM_{2.5}$.

We start by estimating two-stage least squares (2SLS) regressions, described in detail below. We show that the first-stage regressions are not particularly weak, and are unlikely to suffer from omitted variable bias. However, with tests based on Conley, Hansen, and Rossi (2012), we find that the 2SLS estimates are sensitive to violation of the exclusion restriction in our setting. These findings are consistent with prior literature that suggests that biomass burning is associated with pollutants, such as nitrogen oxides and volatile organic compounds, which can also be carried by wind (Andreae, 2019).

We therefore refrain from interpreting 2SLS estimates in the main text. This does not imply that air pollution, broadly defined, is not the mechanism through which agricultural fires affect hypertension risk. Rather, epidemiological and medical literature (e.g., Hadley, Vedanthan, and Fuster, 2018; Cosselman, Navas-Acien, and Kaufman, 2015), as well as the associations that we find in section 2.2, point to that mechanism. Future work should test for the effects of a suite of pollutants, within a unified framework. This is challenging, however, due to the lack of ground-level monitoring of pollutants, especially in agricultural areas of the developing world.

E.1 IV regression specification

To link $PM_{2.5}$ exposure with hypertension risk, one could consider a two-stage least squares (2SLS) regressions, as follows:

First Stage:

$$K_{c,t} = \alpha_1 Fires^{up}_{c,t-1} + \alpha_2 Fires^{down}_{c,t-1} + \alpha_3 Fires^{oth}_{c,t-1} + \boldsymbol{\phi}_x \mathbf{X}_{i,t} + \boldsymbol{\phi}_w \mathbf{W}_{c,t} + \boldsymbol{\theta}_{c,m} + \varepsilon_{c,t} \quad \text{(E.1)}$$

Second Stage:

$$Y_{i,c,t} = \gamma_1 \hat{K}_{c,t} + \gamma_2 Fires^{down}_{c,t-1} + \gamma_3 Fires^{oth}_{c,t-1} + \boldsymbol{\omega}_x \mathbf{X}_{i,t} + \boldsymbol{\omega}_w \mathbf{W}_{c,t} + \boldsymbol{\theta}_{c,m} + \nu_{i,t} \quad \text{(E.2)}$$

where $K_{c,t}$ and $\hat{K}_{c,t}$ are, respectively, observed and predicted $PM_{2.5}$ levels in local cluster c and on the health survey day t; $Y_{i,c,t}$ is the blood pressure outcome for person i; $Fires^{up}_{c,t-1}$ is the number upwind fires in the 24 hours leading to the blood pressure test; $Fires^{down}_{c,t-1}$ are downwind fires; $Fires^{oth}_{c,t-1}$ are fires from other wind directions; and the other variables are as defined in the "reduced form" specification from section B.1 above.

In this approach, the coefficient of interest is γ_1, which identifies how $PM_{2.5}$ exposure affects incidence of high blood pressure. The instrumental variable approach relies on two main assumptions: (a) "exclusion restriction" – the IV (upwind fires) affects the outcome of interest (cardiovascular distress) only through the exposure variable ($PM_{2.5}$); (b) "relevance" – the IV must be associated with the exposure variable.

In Figure 3 from the main text we provide evidence that the assumption "relevance" holds, by showing that the first-stage coefficient is statistically significant. Further, we implement the Kleibergen-Paap "weak-IV" F-test which is robust to non-homoskedastic errors. The resulting test statistic is shown in Table E.1. The F-statistic across all model specifications is above the typical threshold level of 10 used as a cutoff for weak instruments (Andrews, Stock, and Sun, 2019), although recent work in progress recommends using more conservative thresholds (Lee et al., 2021).

The IV estimates remain stable across a variety of specifications presented in Table E.2. First, in columns (1) and (2), we present results from specifications that include a quadratic term of the fire counts (both for upwind fires, the instrument, as well as fires in downwind and other directions used as controls). These estimates are slightly smaller in magnitude, but the confidence intervals overlap with the main estimates obtained from using a linear count of fires. Columns (3) to (6) show estimates from models where we progressively exclude control variables. Columns (3) and (4) drop the individual and household level controls, while columns (5) and (6) also drop the weather variables (temperature, precipitation and wind speed). Across all columns, we see that estimates remain close to those from the

benchmark specification, lending further validity to the relevance assumption.

Finally, we discuss the validity of the "exclusion restriction" in section E.3. We test the sensitivity of the 2SLS estimates to violations of that assumption through a procedure proposed by Conley, Hansen, and Rossi (2012).

E.2 Results from instrumental variable approach

Table E.1: IV estimates – Impact of $PM_{2.5}$ exposure on incidence of hypertension – benchmark specification

	IV using up-wind fires within:	
	75 km	100 km
PM 2.5 ($\mu g/m^3$)	3.834**	3.242**
	(1.790)	(1.615)
'Control' average	93.35	93.35
Observations	783773	783773
1^{st} stage F-stat	15.07	15.92

Notes: Table shows estimates from separate IV-2SLS regressions in each column using 75, and 100 km radius buffers for fire counts. The dependent variable is incidence of hypertension (systolic blood pressure $\geq$ 140 mmHg or diastolic BP$\geq$ 90 mmHg) per '000. The third to last row shows the 'control' average hypertension risk for a sample of individuals that were not exposed to upwind fires. Each column shows the coefficient on 24-hour average $PM_{2.5}$ ($\mu g/m^3$) on the day leading to BP measurement. $PM_{2.5}$ is instrumented using the total number of upwind fires within the specified radius. upwind is based on 90-degree wind sectors. All specifications include district-by-month of sample and day of week fixed effects. Regressions control for weather variables (temperature, rainfall, wind-speed and wind-direction), household and individual characteristics. Standard errors in parentheses are clustered at the NFHS location level. 1^{st} stage F-stat shown is the Kleibergen-Paap F-test statistic. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table E.2: Testing the "relevance" assumption for the IV estimates linking $PM_{2.5}$ exposure and incidence of hypertension

	With quadratic of fire counts		Excluding HH individual controls		Also excluding weather controls	
	(1) 75 km	(2) 100 km	(3) 75 km	(4) 100 km	(5) 75 km	(6) 100 km
PM 2.5 ($\mu g/m^3$)	2.841*	3.138**	4.750**	4.515**	4.391**	3.964**
	(1.523)	(1.497)	(2.059)	(1.924)	(1.910)	(1.695)
'Control' average	93.5	93.5	93.5	93.5	93.5	93.5
Observations	783,773	783,773	803,885	803,885	803,885	803,885
1^{st} stage F-stat	8.42	8.68	13.75	14.40	14.72	16.47

Notes: Table shows estimates from separate IV-2SLS regressions in each column. The dependent variable is incidence of hypertension (systolic blood pressure $\geq$ 140 mmHg or diastolic BP$\geq$ 90 mmHg) per '000. The third to last row shows the 'control' average hypertension risk for a sample of individuals that were not exposed to upwind fires. Each column shows the coefficient on 24-hour average $PM_{2.5}$ ($\mu g/m^3$) on the day leading to BP measurement. In columns (1) and (2) $PM_{2.5}$ is instrumented using the count and the square of the count of upwind fires within the specified radius. We also include quadratic terms for count of downwind and fires in other directions in the controls. Columns (3) and (4) present results from the main linear specification dropping all individual and household level characteristics from the controls. Columns (5) and (6) also drop the weather controls used in the main regression. upwind is based on 90-degree wind sectors. All specifications include district-by-month of sample and day of week fixed effects. Regressions in columns (1) and (2) control for weather variables (temperature, rainfall, wind-speed and wind-direction), household and individual characteristics. Standard errors in parentheses are clustered at the NFHS location level. Significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

E.3 Tests for sensitivity to violations of the IV exclusion restriction

Conley, Hansen, and Rossi (2012) propose methods for inference when the exclusion restriction of Instrumental Variables may not hold. They consider that the instrumental variable, Z, could have a direct effect on the outcome, besides the effect through the endogeneous independent variable of interest X. This is shown in the following system of equations:

$$Y = X\beta + Z\gamma + \varepsilon \tag{E.3}$$
$$X = Z\pi + \nu \tag{E.4}$$

where X is a matrix of endogenous variables, Z is a matrix of instruments, which is expected to be uncorrelated with the error term ε. Equation (E.3) can be thought of as a structural equation which establishes the true relationship between X and Y. Equation (E.4) represents a first-stage regression within the context of IVs. In Equation (E.3), the inclusion of $Z\gamma$ implies that, when $\gamma \neq 0$, the exclusion restriction assumption does not hold. The proposed inference approach relies on relaxing the exclusion restriction, such that γ need not be exactly zero, although it should be near to zero. In this setting, Conley, Hansen, and Rossi (2012) derive the potential bias of the 2SLS estimator of β as follows:

$$\hat{\beta}^{2SLS} = \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}\left(X'Z(Z'Z)^{-1}Z'Y\right)$$

Then, replacing Y from Equation (E.3):

$$\begin{aligned}\hat{\beta}^{2SLS} = &\left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}\left(X'Z(Z'Z)^{-1}Z'X\right)\beta + \\ &\left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}\left(X'Z(Z'Z)^{-1}Z'Z\right)\gamma + \\ &\left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}\left(X'Z(Z'Z)^{-1}Z'\varepsilon\right)\end{aligned}$$

$$\begin{aligned}\hat{\beta}^{2SLS} = &\beta + \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}\left(X'Z\right)\gamma + \\ &\left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}\left(X'Z(Z'Z)^{-1}Z'\varepsilon\right)\end{aligned}$$

$$\text{Now let } A \equiv \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}\left(X'Z\right)$$
$$\hat{\beta}^{2SLS} = \beta + A\gamma + A(Z'Z)^{-1}Z'\varepsilon$$
$$\hat{\beta}^{2SLS} - \beta = A\gamma + A(Z'Z)^{-1}Z'\varepsilon$$

Since Z is assumed to be uncorrelated with ε, then $A(Z'Z)^{-1}Z'\varepsilon \xrightarrow{p} 0$. Therefore, the asymptotic bias is given by $A\gamma$. Under the exclusion restriction, $\gamma = 0$ and $\hat{\beta}^{2SLS} \xrightarrow{p} \beta$, such that the 2SLS estimator is consistent. However, if $\gamma \neq 0$, the asymptotic bias will depend on the sizes of γ and A. As shown in Conley, Hansen, and Rossi (2012), stronger instruments are associated with smaller A. This derivation thus highlights how deviations from the exclusion restriction can generate significant bias under weak instruments.

To account for this, Conley, Hansen, and Rossi (2012) propose an inference procedure that is based on the asymptotic distribution of $\hat{\beta}^{2SLS}$. Under the Central Limit Theorem, and assuming $\gamma = 0$, we have:

$$\sqrt{N}(\hat{\beta}^{2SLS} - \beta) \xrightarrow{d} \mathcal{N}(0, A(Z'Z)^{-1}Z'\varepsilon^2 Z(Z'Z)^{-1}A)$$

such that:

$$(\hat{\beta}^{2SLS} - \beta) \xrightarrow{d} \mathcal{N}(0, Var(\beta))$$

where $Var(\beta)$ is the variance-covariance matrix of β.

For $\gamma \neq 0$, Conley, Hansen, and Rossi (2012) propose a bootstrap procedure to obtain B draws of the 2SLS bias $\Delta = \hat{\beta}^{2SLS} - \beta$. Empirically, each draw of Δ will be composed of a draw from a random distribution

$\mathcal{N}(0, \hat{Var}(\hat{\beta}^{2SLS}))$, plus a draw from a distribution of γ multiplied by A, as follows

$$\Delta \approx \mathcal{N}(0, \hat{Var}(\hat{\beta}^{2SLS})) + A\gamma$$

Note that $\hat{Var}(\hat{\beta}^{2SLS})$ and A can be estimated, while γ is an unknown parameter. The distribution of γ needs to be set by the researcher, based on prior knowledge about the specific setting of interest. Finally, it is possible to construct a $(1-\alpha)\%$ confidence interval based on the percentiles $\alpha/2$ and $1-\alpha/2$ of the distribution of Δ from the B random draws, as follows:

$$\left(\hat{\beta}^{2SLS} - c_{1-\alpha/2}; \hat{\beta}^{2SLS} - c_{\alpha/2}\right) \tag{E.5}$$

However, in many settings, Z is unlikely to have a direct impact on Y. Rather, we might expect some omitted variable W (e.g., other pollutants) that affect Y and are also related to the instrument Z, as follows:

$$Y = X\beta + W\lambda + \varepsilon \tag{E.6}$$
$$X = Z\pi_X + \nu_X \tag{E.7}$$
$$W = Z\pi_W + \nu_W \tag{E.8}$$

In such cases, λ captures the direct effect of W on Y, while π_W captures the correlation between the omitted pollutant W and the instrument Z. The bias of the 2SLS estimator for β will then be:

$$\hat{\beta}^{2SLS} = \beta + \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}\left(X'Z(Z'Z)^{-1}Z'W\right)\lambda + \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}\left(X'Z(Z'Z)^{-1}Z'\varepsilon\right)$$

Note that $(Z'Z)^{-1}Z'W \equiv \pi_W$

and $A \equiv \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}(X'Z)$, such that:

$$\hat{\beta}^{2SLS} = \beta + A\pi_W\lambda + A(Z'Z)^{-1}Z'\varepsilon$$
$$\hat{\beta}^{2SLS} - \beta = A\pi_W\lambda + A(Z'Z)^{-1}Z'\varepsilon$$

In this case, the size of the potential bias depends on A, λ, and π_W, such that:

$$\Delta \approx \mathcal{N}(0, \hat{Var}(\hat{\beta}^{2SLS})) + A\pi_W\lambda \tag{E.9}$$

If either λ or π_W are equal to zero, then the exclusion restriction holds. However, if the instrument is correlated with the other pollutants ($\pi_W \neq 0$), and if the other pollutants have an impact on health ($\lambda \neq 0$), then $\hat{\beta}^{2SLS}$ will be biased. Following the bootstrap procedure from Conley, Hansen, and Rossi (2012), described above, we can adjust confidence intervals for $\hat{\beta}^{2SLS}$ that account for this potential bias.

For this procedure, we need to draw from distributions of both λ and π_W. The objective is to provide approximate bounds on λ and π_W, which would make $\hat{\beta}^{2SLS}$ not statistically significant. As an illustrative example, we assume that nitrogen oxides (NO_X) are the only pollutants, other than $PM_{2.5}$, that are associated with upwind fires, and that can affect hypertension risk. For a starting point on π_W, we take estimates from Andreae (2019) which suggest that the biomass burning emissions factor for NO_X approximately one third that for $PM_{2.5}$. Based on this, we run scenarios with $\pi_W \sim \mathcal{N}(r\hat{\pi_X}, r^2\hat{Var}(\hat{\pi_X}))$, where r is a rescaling factor equal to 0.2, 0.3, or 0.4, and $\hat{\pi_X}$ is our first stage estimate of the effects of upwind fires on $PM_{2.5}$. For the simulated effect of NO_X on hypertension, we consider scenarios with $\lambda \sim \mathcal{N}(r\hat{\beta}^{2SLS}, r^2\hat{Var}(\hat{\beta}^{2SLS}))$, and with $r = 0.2$, 0.4, 0.6, 0.8, or 1. For each scenario, we perform one thousand bootstrap draws from these distributions.

Simulation results are in Table E.3, which presents the adjusted 95% confidence intervals of $\hat{\beta}^{2SLS}$ across all scenarios. We find that $\hat{\beta}^{2SLS}$ remains statistically significant only if the effects of other pollutants on hypertension are relatively small (i.e., $\lambda \leq 0.2\hat{\beta}^{2SLS}$ and $\pi_W \leq 0.4\hat{\pi_X}$). In particular, we note that a moderate first stage relationship ($\pi_W > 0.4\hat{\pi_X}$) between upwind fires and other pollutants is sufficient to cast doubt on the consistency of $\hat{\beta}^{2SLS}$. These findings, therefore, caution against a causal interpretation of the 2SLS estimates from section E.2.

Table E.3: Sensitivity to violations of the exclusion restriction

	$\pi_W \sim 0.2 \times \hat{\pi_X}$		$\pi_W \sim 0.3 \times \hat{\pi_X}$		$\pi_W \sim 0.4 \times \hat{\pi_X}$	
	LB	UB	LB	UB	LB	UB
$\lambda \sim 0.2 \times \hat{\beta}^{2SLS}$	[0.324	7.180]	[0.200	7.256]	[0.038	6.814]
$\lambda \sim 0.4 \times \hat{\beta}^{2SLS}$	[0.020	6.945]	[-0.467	7.034]	[-0.324	6.765]
$\lambda \sim 0.6 \times \hat{\beta}^{2SLS}$	[-0.210	7.051]	[-0.522	6.851]	[-0.888	6.644]
$\lambda \sim 0.8 \times \hat{\beta}^{2SLS}$	[-0.516	7.024]	[-0.688	6.613]	[-1.088	6.609]
$\lambda \sim \hat{\beta}^{2SLS}$	[-0.284	6.723]	[-1.127	6.467]	[-1.366	6.082]

Notes: This table presents 95% confidence intervals for the 2SLS estimate of the effect of $PM_{2.5}$ on hypertension risk, adjusted based on a proposal from Conley, Hansen, and Rossi (2012). The size of the potential bias is estimated from one thousand bootstrap draws of Δ from equation E.9. We assume that the exclusion restriction may be violated due to other pollutants W that can affect hypertension, with a first stage coefficient of π_W and a "true" effect on hypertension of λ. "LB" refers to the lower bound effect, while "UB" refers to the upper bound.