How Do Household Energy Transitions Work?

Jill Baumgartner (Co-PI) Sam Harper (Co-PI)
On behalf of the Beijing Household Energy Transitions Team

2024-04-24

Table of contents

Abstract

Introduction

Methods

Results

Conclusions

1 Introduction

China is deploying an ambitious policy to transition up to 70% of households in northern China to clean space heating, including a large-scale roll out across rural and peri-urban Beijing, referred to in this document as China's Coal Ban and Heat Pump (CBHP) subsidy policy. To meet this target the Beijing municipal government announced a two-pronged program that designates coal-restricted areas and simultaneously offers subsidies to night-time electricity rates and for the purchase and installation of electric-powered, air-source heat pumps to replace traditional coal-heating stoves. The policy was piloted in 2015 and, starting in 2016, was rolled out on a village-by-village basis. The variability in when the policy is applied to each village allows us to treat the roll-out of the program as a quasi-randomized intervention. Households may also be differentially affected by this program due to factors such as financial constraints, preferences and social capital, and there is uncertainty about whether and how this intervention may affect indoor and outdoor air pollution, as well as heating behaviors and health outcomes.

2 Background

2.1 Context for the policy

Beijing has a temperate continental monsoon climate characterized by cold, dry winters and hot, humid summers. Access to central heating is limited to urban areas and households in most rural and peri-urban areas of Beijing have historically heated their homes using coal heaters and biomass-fueled *kangs* (a traditional Chinese energy technology that integrates at least four different home functions including cooking, a bed for sleeping, space heating, and home ventilation). Household coal burning was a major contributor to indoor and outdoor air pollution in northern China, especially in winter. In 2015, over 100 million rural households consumed around 200 million tons of coal to meet over 80% of northern China's residential space heating demand (Dispersed Coal Management Research Group 2023). At that time, household coal-fuelled heaters burned approximately half of the over 400 million tons of coal used for space heating (Group 2016) and contributed to roughly 30% of northern China's wintertime air pollution. In 2013, exposure to ambient fine particulate matter from coal combustion - from industry, electricity, and domestic sources - was the largest estimated contributor to population exposure to $PM_{2.5}$ and contributed to an estimated 366,000 premature deaths annually in China (Group 2016).

Banning residential coal burning and replacing household coal stoves with clean heating alternatives was considered a potentially impactful intervention to improve rural development, reduce $PM_{2.5}$ across the region, and mitigate air pollution-related health impacts. A number of clean heating options, including electric heat pumps, gas heaters, pelletized biomass stoves, and electric resistance heaters with thermal storage, were widely promoted by the Chinese government (Dispersed Coal Management Research Group 2023). By 2021, over 36 million households in northern China were treated by the policy and an estimated 21 million additional households are expected to be treated by 2025. Whether this large-scale energy policy yielded air quality and health benefits remains a critical and unresolved question.

2.2 Prior evidence on household energy interventions and air pollution

Household energy interventions, mostly cooking-related, that replace solid fuel stoves with cleaner-burning alternatives have been implemented and studied extensively in countries including China over the past several decades. While their introduction of more efficient residential stoves and fuels is expected to reduce air pollution emissions and subsequent exposures, there is still no consensus about their effectiveness in achieving health-relevant air pollution reductions (Quansah et al. 2017). In particular, the effectiveness of large-scale household energy programs like China's Coal Ban and Heat Pump (CBHP) subsidy policy have been rarely empirically investigated, especially at sub-city spatial resolution. In Ireland, county-level residential coal bans in the 1990s were associated with 40-70% decreases in black smoke concentrations in ban-affected areas (Dockery et al. 2013). In

Australia, a wood-burning stove exchange lowered daily wintertime PM_{10} from 44 to 27 µg/m^3 (Johnston et al. 2013), and clean energy policies in New Zealand were associated with 11-36% reductions in winter PM_{10} (Scott and Scarrott 2011). The few evaluations of the CBHP policy reported small decreases in outdoor $PM_{2.5}$ (-7 to -2.4 µg/m^3) in municipalities or prefectures in the policy compared with neighboring areas not affected by the policy (Niu et al. 2024; Song et al. 2023; Tan et al. 2023; Yu et al. 2021), and a recent modeling study estimated 36% lower personal exposure to $PM_{2.5}$ based on household-reported changes in fuel use (Meng et al. 2023). However, none of these studies included field-based measurements of air pollution or personal exposures, which are known to differ considerably from modeled estimates (Thompson et al. 2019), and few accounted for secular changes in air quality over time, limiting any conclusions about the causal effect of the CBHP policy on air quality.

2.3 Prior evidence on clean energy interventions and cardiovascular outcomes

Most previous health assessments of household energy interventions have focused on cookstoves instead of heating, and randomized trials of less polluting cookstoves generally indicate a cardiovascular benefit. In older Guatemalan women, a chimney stove intervention lowered exposure to air pollution and reduced the occurrence of nonspecific ST-segment depression (McCracken et al. 2011). Randomized trials in Guatemala, Nigeria, and Ghana also showed reductions in blood pressure (systolic range: -3.7 to -1.3 mmHg) in women assigned to gas, ethanol, or improved combustion biomass stoves. In contrast, a recent multi-country randomized trial found little evidence for a protective effect of gas stoves on gestational blood pressure (Ye et al. 2022) despite large reductions (~66% lower) in exposure to $PM_{2.5}$(Johnson et al. 2022).

The few population-based evaluations of large-scale residential energy policies also suggest a cardiorespiratory benefit of clean energy transition. Residential wood-burning bans were associated with reductions in cardiovascular hospitalizations (-7%) in California (Yap and Garcia 2015) and with reduced cardiovascular (-17.9%) and respiratory (-22.8%) mortality in Australia (Johnston et al. 2013), though neither study fully controlled for possible secular improvements in health that were unrelated to the policy. Most relevant to our study are two quasi-experimental assessments of coal replacement policies. In Ireland, reductions in respiratory not but cardiovascular mortality were observed following a coal ban (Dockery et al. 2013). A multi-city study of Chinese adults in cities where the CBHP policy was piloted compared with adults in cities not in the pilot observed small decreases in chronic lung diseases (-3.0 to -1.1%) but no change in physician-diagnosed cardiovascular diseases, potentially due to the short (one-year) post-policy evaluation period or confounding by other unmeasured city-wide air quality or health-related policies (Wen et al. 2023).

Though household air pollution is a well-established health risk factor, which energy interventions can reduce air pollution exposures, improve health, and are scalable and sustainable remains a critical and unanswered question. In a recent Official American Thoracic Society Statement, for example, the committee could not reach a consensus on whether previously studied household energy interventions (including gas, ethanol, solar, and improved biomass stoves) improved health outcomes that included blood pressure (with 55% saying no and 45% saying yes) (Harrison et al. Approved February 2024).

2.4 Assessing dynamic and heterogeneous treatment effects

Since 2015, thousands of villages across Beijing and northern China entered the CBHP policy prior to the start of the heating season each year. Given the many behavioral, social, or economic factors that might affect both new heater use and coal stove suspension (e.g., energy prices and availability, wintertime temperature, COVID-19 pandemic, user preferences), it is possible that the effect of the policy on air pollution and health may be dynamic over time and/or heterogeneous across treatment cohorts. Thus, it may be important to study both the overall and group-time effects of the policy.

2.5 Evaluating the mechanisms through which policies may affect health outcomes.

With several exceptions (Alexander et al. 2018; Gould et al. 2023; McCracken et al. 2007; McCracken et al. 2011), decades of household energy intervention studies showing limited or no health benefit demonstrate the complexity of evaluating interventions on exposures such as cooking or space heating that are central to daily life (Ezzati and Baumgartner 2017) (Harrison et al. Approved February 2024). Energy interventions and policies, particularly those implemented at the household- or village-scales, can produce multiple behavioral, environmental, and health-related changes, making it important to investigate the mechanisms through which such policies exert their health impacts (Dominici et al. 2014). The health benefits achievable with transition from traditional coal stoves to a new electric home heating system, for example, may be influenced by factors including outdoor air quality (Lai 2019), the desirability and usage patterns of new and traditional stoves (Ezzati and Baumgartner 2017), indoor temperature (Lewington et al. 2012), or behaviors including physical activity (Lindemann et al. 2017). Only recently were these mediating factors considered in health assessments of household energy interventions, and rarely in a comprehensive or formalized way (Rosenthal et al. 2018). Understanding these mechanisms can provide valuable insights into the success (or failure) of clean energy programs or policies like the CBHP policy in meeting their air quality and health goals, and may answer questions that can inform the design of more effective future energy interventions (Harrison et al. Approved February 2024). For example, is there successful uptake of the intervention or policy? Does the policy lead to heating behavior changes that result in colder homes and thus offsets any cardiovascular-enhancing effects of improved air quality? Answers to these questions are facilitated by the analysis of mediating pathways.

3 Specific Aims and Overarching Approach

This study used three data collection campaigns in winter 2018/19, winter 2019/20, and winter 2021/22, as well as a partial campaign in winter 2020/21 to advance the following aims:

1. Estimate how much of the CBHP policy's overall effect on health, including respiratory symptoms and cardiovascular outcomes (blood pressure, blood inflammatory and oxidative stress markers), can be attributed to its impact on changes in $PM_{2.5}$;
2. Quantify the impact of the policy on outdoor air quality and personal air pollution exposures, and specifically the source contribution from household coal burning;
3. Quantify the contribution of changes in the chemical composition of $PM_{2.5}$ from different sources to the overall effect on health outcomes.

4 Study Design and Methods

4.1 Study area

Beijing is the capital of China (pop. 21.9 million in 2020) and covers a large geographic area (~16,000 km2) that includes a highly developed and densely-populated urban core that is surrounded by several satellite towns and peri-urban and rural villages in the periphery. Beijing winters begin in early November and tend to be cold, dry, and windy with the lowest temperatures mostly often occurring in January (-3°C, on average), thus requiring space heating (An et al. 2021). Most urban areas of Beijing are connected to a central heating grid that supplies home heating from central locations, whereas rural and many peri-urban areas have historically relied on individual space heating units that, prior to 2015, were largely fueled by unprocessed raw coal (Duan et al. 2014).

4.2 Location and participant recruitment and enrolment

Between December 2018 and January 2019 we recruited 50 villages across 4 administrative districts (Fangshan, Huairou, Mentougou, and Miyun) in the Beijing municipality in northern China. The villages predominately used coal for heating at the time of enrollment and were eligible for and not currently participating in the CBHP policy. Roughly half of the villages were expected to enter into the policy during our study (Figure 1). We used local guides in each village to help determine a roster of households that were not vacant during the winter months, from which we randomly selected households to recruit for participation.

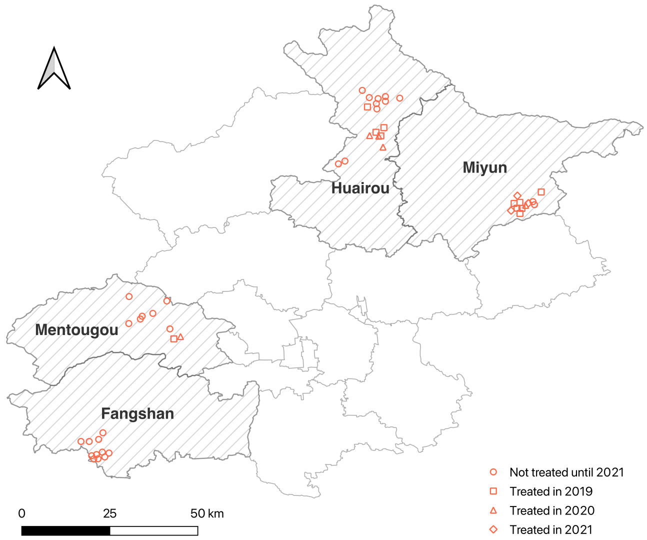


Figure 1: Map of village implementation of CBHP policy

We recruited approximately 20 households in each village and randomly selected one eligible person from each household to participate. Household members were eligible to participate if they were over 40 years old, lived in the study villages, were not planning to move out of the village in the next year, and were not on current immunotherapy or treatment with corticosteroids. Research staff introduced the study and its measurements to an eligible adult in each household and answered any questions related to the study. In follow-up visits to the study villages, staff first approached households with participants from an earlier campaign. If previous participants were not at home or refused to participate, staff first tried to randomly recruit an eligible participant from the same household. If there was not another eligible or willing participant in the household, we randomly selected and recruited a participant from a new household using the village roster. All participants provided written informed consent prior to joining the study. The study protocols were approved by research ethics boards at Peking University (IRB00001052-18090), Peking Union Medical College Hospital (HS-3184) and McGill University (A08-E53-18B).

4.3 Data Collection Overview

We conducted study measurements over four consecutive winter seasons in 2018-19, 2019-20, 2020-21, and 2021-22 (referred to hereafter as Season 1 [S1], S2, S3 and S4, respectively). Field data collection was conducted by ~20 trained staff members who traveled to participants' homes to conduct tablet-based household and individual questionnaires, measure participant blood pressure, and distribute temperature sensors (for measurement of indoor temperature and stove use) and air pollution monitors in all 50 study villages in S1, S2, and S4. Anthropometrics (height, weight, and waist circumference), measurement of airway inflammation, and whole blood samples were obtained no more than a month later at a village clinic in S1 and S2. In S3, which was during the height of the pandemic, we limited household measurements to indoor air quality and sensor-based measurement of indoor temperature and stove use in 41 villages, including all 17 treated villages and 24 untreated (control) villages, prior to COVID-related travel restrictions that halted field data collection. In S4, which also occurred during the COVID-19 pandemic, we returned to conducting individual-level assessments. However, unlike in S1 and S2, anthropometric measurements and airway inflammation were assessed in participant homes rather than clinics to avoid group contact and blood samples were not collected. Outdoor (community) air pollution was measured throughout the study period.

4.3.1 Air Pollution

Outdoor air pollution

In each village, two sensors for particulate matter air pollution were set up to measure outdoor (community) $PM_{2.5}$ at different locations in each village. One sensor was placed near the center of the village, and the other was placed no less than 500m away from the centrally-located sensor. Sensors were placed at least 1.5m above the ground and not in a location within sight of a visible point source of $PM_{2.5}$.

We collected filter-based community $PM_{2.5}$ samples to calibrate the sensor-based $PM_{2.5}$ measurements as well as to conduct analysis of chemical composition for source apportionment. Ultrasonic Personal Aerosol Samplers (UPAS, Access Sensor Technologies, Fort Collins, CO, USA) were used to collect filter-based $PM_{2.5}$ samples with a flow rate of 1.0 L/min (Volckens et al. 2017). Samplers housed 37mm PTFE filters (VWR, 2.0- m pore size) and were equipped with a cyclone inlet with a 2.5 m cut point designed to perform under the sampling flow rate. For community measurements, a UPAS was co-located with each $PM_{2.5}$ sensor in each village in rotation. Every week, the used filters were removed and replaced with a new filter. In total, we successfully collected 126, 371, and 289 filter-based, community outdoor $PM_{2.5}$ samples in Seasons 1, 2, and 4, respectively. Field blank filters were collected at a rate of ~10%, subject to the same field conditions as samples. To support post-sampling determination of organic carbon (OC) and elemental carbon (EC) fractions of $PM_{2.5}$ mass, quartz filters were co-located with a subset of Teflon filter samples collected outdoors. Quartz filter-based $PM_{2.5}$ samples were collected using UPAS operating with a flow rate of 1.0 L/min. UPASs housed 37 mm quartz filters (VWR, 2.0- m pore size) and were equipped with a cyclone inlet with a 2.5 m cut point designed to perform under the corresponding sampling flow rate. All quartz fiber filters were baked at 550 °C for a minimum of 8 h to remove organic impurities prior to sample collection. $PM_{2.5}$ samples collected on quartz filters were analyzed using established thermo-optical methods for quantifying elemental carbon (EC) and organic carbon (OC) to, then, calibrate the colorimetric analysis of EC and OC on Teflon filters. In Season 2, 23 quartz-based outdoor $PM_{2.5}$ samples and 3 field blanks were collected. In Season 4, 11 quartz-based outdoor $PM_{2.5}$ samples and 3 field blanks were collected.

For $PM_{2.5}$ sensor calibration and quality control, all PM sensors were co-located with a reference-grade $PM_{2.5}$ instrument (Model 5030 Synchronized Hybrid Ambient Realtime Particulate (SHARP) Monitor, Thermo Fisher Scientific, United States) on the rooftop of a building at Peking University campus and/or the Tapered Element Oscillating Microbalance (TEOM, Thermo Scientific™ 1405 TEOM™) at the Chinese Academy of Sciences University campus for 7 to 10 days before and after each field campaign (Figure 2). Sensor-measured $PM_{2.5}$ concentrations were highly correlated with those measured by the reference instruments (Spearman correlation coefficients (rho) >0.75 in each pre- and post-calibration).

Figure 2: Calibration of real-time sensors against a reference monitor at University of the Chinese Academy of Sciences.

Indoor $PM_{2.5}$

In the second and fourth field seasons (i.e., S2 and S4), we randomly selected six households from the 20 recruited in each village to measure indoor concentrations of $PM_{2.5}$. In Season 4, we aimed to monitor indoor $PM_{2.5}$ in the same households where we measured indoor $PM_{2.5}$ in Season 2. If a household dropped out of the project or declined indoor $PM_{2.5}$ monitoring, we then

Table 1: Household recruitment for overall and indoor air quality measurements.

	Overall			Indoor		
Sample	Season 1	Season 2	Season 4	Season 1	Season 2	Season 4
New recruitment	977	196	68	0	300	52
Households from Season 1	\	866	780	\	0	0
Households from Season 2	\	\	162	\	\	248
Total recruitment	977	1062	1010	0	300	300

recruited another household already enrolled in this study to measure indoor $PM_{2.5}$. In total, indoor measurements were conducted in 300 households in both Season 2 and Season 4 (Table 1).

Time-resolved indoor $PM_{2.5}$ concentrations were measured using the same commercially available sensor (PMS7003 Plantower, Zefan, Inc.) as was used for outdoor sensor-based $PM_{2.5}$ measurements and recorded $PM_{2.5}$ concentrations every 1 min. The sensor was placed on a table in a room where participants reported spending most of their time when awake, e.g., a living room or bedroom. Indoor $PM_{2.5}$ sensors were deployed between late November and mid January within field seasons (i.e., S2 and S4), depending on the village and household visit schedule. The measurement continued from the time of deployment until sensors were recollected from homes in late April to capture the full heating season.

We randomly selected three households from the six in which we deployed $PM_{2.5}$ sensors to co-locate a filter-based $PM_{2.5}$ sampler with the $PM_{2.5}$ sensor. We collected a 24-h $PM_{2.5}$ filter sample at the first 24-h of indoor $PM_{2.5}$ sensor measurements. Filter-based $PM_{2.5}$ samples were collected using Ultrasonic Personal Aerosol Samplers (UPAS, Access Sensor Technologies) or Personal Exposure Monitors (PEMs, Apex Pro) operating with flow rates of 1.0 and 1.8 L/min, respectively. Both samplers housed 37 mm PTFE filters (VWR, 2.0- m pore size) and were equipped with a cyclone inlet with a 2.5 m cut point designed to perform under the corresponding sampling flow rate. After 24-h, the samplers were retrieved and loaded with new filters for measurements in other villages, once the previous sample filters were removed and stored for later analysis. In total, we successfully collected 149 and 148 indoor $PM_{2.5}$ filter samples in S2 and S4, respectively.

As with the community outdoor air sampling, to support post-sampling determination of organic carbon (OC) and elemental carbon (EC) fractions of $PM_{2.5}$ mass, quartz filters were co-located with a subset of Teflon filter samples collected in homes. Filter-based $PM_{2.5}$ samples were collected using Personal Exposure Monitors (PEMs, Apex Pro) operating with flow rates of 1.8 L/min. PEMs housed 37 mm quartz filters (VWR, 2.0- m pore size) and were equipped with a cyclone inlet with a 2.5 m cut point designed to perform under the corresponding sampling flow rate. All quartz fiber filters were baked at 550 °C for a minimum of 8 h to remove organic impurities prior to sample collection. $PM_{2.5}$ samples collected on quartz filters were analyzed using established thermo-optical methods for quantifying elemental carbon (EC) and organic carbon (OC) to, then, calibrate the colorimetric analysis of EC and OC on Teflon filters. In Season 2, 71 quartz-based indoor $PM_{2.5}$ samples and 14 field blanks were successfully collected. In Season 4, indoor $PM_{2.5}$ samples for gravimetric analysis had to be collected on two types of PTFE sample media (Zefluor and Teflo filters), due to discontinuation of manufacturing of the Zefluor filter media. To ensure that quartz filters were deployed with both types of Teflon-based filter media, 73 quartz-based indoor $PM_{2.5}$ samples were collected concurrently with Zefluor samples, and 47 quartz indoor $PM_{2.5}$ samples were collected alongside Teflo samples. For indoor quartz $PM_{2.5}$ mass sampling in S4, 18 field blanks were collected.

Personal exposure to $PM_{2.5}$ and black carbon

To measure personal exposure we used two types of samplers: Personal Exposure Monitors (PEMs, Apex Pro; Casella, UK) and Ultrasonic Personal Aerosol Samplers (UPAS, Access Sensor Technologies, Fort Collins, CO, USA). PEMs actively sampled air at a flow rate of 1.8 L/min, and UPAS sampled air at 1.0 L/min (Volckens et al. 2017). Both samplers housed 37 mm PTFE filters (VWR, 2.0- m pore size) and were equipped with a cyclone inlet with a 2.5 m cutpoint. Sampler flow rates were calibrated the night before deployment and measured immediately after the sampling period. Only 2% of the post-sampling measurements deviated from the target flow rate by greater than +/-10%. Participants were instructed to wear a small waistpack (for the PEM and sampling pump) or an arm band or cross-body sling (for the UPAS) for 24 hours, which they could remove from their body and place within 2 meters while sleeping, sitting, or bathing. Field blanks for personal air pollution exposure measurements were collected at a rate of ~10% in each village.

Gravimetric analyses of PTFE filter-based $PM_{2.5}$ samples

All filters were placed in individually labeled cases, sealed in plastic bags, and then transported to a field laboratory and immediately stored in a -20°C freezer. Following completion of the field sampling campaign, the samples and blanks were transported to Colorado State University, where they were stored in a -20°C freezer prior to gravimetric and chemical analysis of $PM_{2.5}$.

All filters were placed in an environmentally-controlled equilibration chamber (21-22 °C, 30-34% relative humidity) for at least 24 hours before tare and gross weighing. Before each weight was taken, filters were discharged by a polonium-210 strip. Filters were weighed on a microbalance (Mettler Toledo Inc., XS3DU, USA) with 1- g resolution in triplicate or more, until the differences among three weights were less than 3 g. The average of three readings was used to determine filter mass, which was then blank-corrected using the median value of blank filters [3 g for UPAS-collected filters (53% of samples); 33 g for PEM-collected filters (47% of filter samples)], and $PM_{2.5}$ concentrations were calculated by dividing the mass by the sampled air volume.

Adjusting sensor-based $PM_{2.5}$ using filter-based gravimetric measurements

We established linear regression models between the filter-based $PM_{2.5}$ mass concentrations (i.e., the 'gold standard' reference concentrations) and the sensor-based $PM_{2.5}$ concentrations averaged over the same sampling period as the filter-based samples. The slopes of the models were used as the adjustment factors for the sensor-based $PM_{2.5}$ concentrations. Separate regression models were conducted for indoor and outdoor sensors and for each season given the sensitivity of the sensors to relative humidity, temperature, and particle sources, which may differ for indoor versus outdoor conditions and across seasons. In Season 3, where only sensor-based measurements were conducted for indoor $PM_{2.5}$ to avoid direct contact with household members during the COVID-19 pandemic, we applied an adjustment factor developed from a linear regression model that incorporated data from both Season 2 and Season 4.

The PM sensors were also evaluated before and after each season to identify any sensors that needed further repair or replacement. The $PM_{2.5}$ sensors underwent a calibration process that began with synchronization to real-time $PM_{2.5}$ monitors at Peking University (PKU) campus. This pre- and post-season calibration included a week-long session using the Beta Attenuation Monitor (BAM) alongside daily 24-hour filter samples. During this time, approximately 240 sensors were placed on the rooftop of the College of Urban and Environmental Sciences building, each recording data every minute. A similar approach was taken at the University of Chinese Academy of Sciences (UCAS) campus, where around 400 PM sensors were installed on the rooftop of the Environmental Monitoring Site of the College of Resources and Environment, with data logging at one-minute intervals. Daily collections of 24-hour Zeflour (Teflon) and quartz filter samples accompanied the sensors' measurements to ensure accuracy. The calibration process was repeated post-fieldwork to account for any potential shifts or discrepancies in sensor performance. This approach aimed to maintain consistent and accurate measurements from the PM sensors throughout the study.

Chemical analysis of PM mass

We analyzed the chemical composition of community outdoor and personal exposure $PM_{2.5}$ samples from each season to quantify the individual components and species. $PM_{2.5}$ component concentrations were determined within each community by dividing the quantified component mass by the sampled air volume, after correcting for field blanks collected in the corresponding season.

Elemental analysis of $PM_{2.5}$ mass was performed using a Thermo Scientific Quant'X Evo energy-dispersive X-ray fluorescence (EDXRF) spectrometer with Wintrace software version 10.3 using standard methods (International 2009). Quantitative mass concentrations of 22 individual elements (Mg, Al, Si, S, K, Ca, Ti, Cr, Mn, Fe, Ni, Cu, Zn, Ga, As, Se, Cd, In, Sn, Sb, Te, I) were determined empirically using linear standard curves. Standard curves were generated from commercial, single and dual element, thin film standards from MicroMatter Technologies Inc. (Montreal, Canada) in addition to blank films. The quality of the analysis method was evaluated by analyzing a National Institute of Standards and Technology (NIST) standard reference material (SRM) 2783

Air particulate on filter media (Gaithersburg, MD, USA). Elements for which at least 80% of $PM_{2.5}$ mass samples yielded quantifiable element mass were included for positive matrix factorization and source analysis and apportionment. Those elements were: Si, Mg, Fe, S, Ca, Al, K, Pb.

For analysis of water-soluble ions, a portion of each PTFE filter was extracted in 15 mL deionized water (DI Water) in a Nalgene Amber HDPE bottle using sonication without heat for 40 min. The extracts were filtered to ensure that insoluble particles were removed using a 0.2 m PTFE syringe filter. Water-soluble ions were measured using a dual channel Dionex ICS-3000 ion chromatography system. Specifically, a Dionex IonPac CS12A analytical (3 × 150 mm) column with eluent of 20 mM methanesulfonic acid at a flow rate of 0.5 mL/min was used to measure cations (Ca2+, Mg2+, Na+, NH4+, K+), while a Dionex IonPac AS14A analytical (4 × 250 mm) column with an eluent of 1 mM sodium bicarbonate/8 mM sodium carbonate at a flow rate of 1 mL/min was used to measure anions (SO42−, NO3−, Cl−) (Sullivan et al., 2008).

Organic (OC) and elemental carbon (EC) on PTFE filters were measured using an optical color space sensing system. The CIE-Lab color space optical sensing system measures the optical properties of the $PM_{2.5}$ samples, and these properties are used to develop the EC and OC predictive models. The CIE-Lab color system is a color-opponent space that includes all of the color models, with dimension L* for lightness and a* and b* for the color-opponent dimensions. More information about the CIE Lab color space system, its formulation, and its specific application to the analysis of OC and EC fractions of fine particulate matter pollution is provided in Khuzestani et al. (Khuzestani et al. 2017). Briefly, all the Teflon (PTFE) and quartz filters collected were analyzed using the i1Pro Colorimeter (X-Rite, INC. Grand Rapids, MI). The colorimeter sensor was placed directly over the quartz and Teflon filters, and the color components were measured under the D65 instrument internal illumination light source. Each filter sample was analyzed in triplicate, and the average value of each color coordinate was applied as the optical property of the sample (Olson et al. 2016). CIE Standard Illuminant D65 simulates average midday light and is a commonly used standard illuminant, as defined by the International Commission on Illumination (CIE). The CIE-Lab color space response variables were used in separate random forest models for EC and OC.

The reference measurements for the random forest model development were EC and OC determined from quartz filters collected indoors and outdoors (as described above). $PM_{2.5}$ samples collected on quartz filters were analyzed for OC and EC using a Sunset Laboratory OC/EC Lab instrument (Sunset Laboratories, Inc., MODEL, USA) according to the default Sunset Analyzer protocol. A section of each quartz filter underwent a combined thermal desorption-optical transmittance measurement based on NIOSH methods 5040 to differentiate and quantify the EC and OC components in mass. For the thermal desorption component, the sample is oxidized twice, according to a strict temperature regime. The first oxidation stage thermally removes OC in a mobile phase of pure helium gas to be converted from carbon dioxide (CO2) to methane (CH4) gas and measured by a flame ionization detector (FID). The second oxidation stage proceeds in a mixture of helium and oxygen to oxidize EC, which is also quantified by the FID. The FID is internally calibrated with methane, and external quality control checks are made with sucrose standards. To correct for the potential production of EC by OC pyrolysis during the first heating stage, light transmission from a laser through the filter section was monitored throughout analysis. Reduced light transmittance corresponds to EC generated by the laboratory analysis.

Following gravimetric analysis, all PTFE filters were also analyzed for black carbon (BC) using an optical transmissometer data acquisition system (SootScan^TM OT21 Optical Transmissometer; Magee Scientific, Berkeley, CA, USA). Light attenuation through each filter was measured before and after sampling in the field. To calculate BC mass, the difference between the pre- and post- light attenuation was converted to a mass surface loading using the classical Magee mass absorption cross-sections of 16.6 m2/g for the 880 nm channel optical BC (Ahmed et al. 2009). BC concentrations were calculated by multiplying surface loadings by the sampled surface area of the filters (8.6 cm2 for UPAS-collected filters; 7.1 cm2 for PEM-collected filters), correcting for the field blank mass using the median value of blanks (0.31 g for UPAS-collected filters; 0.01 g for PEM-collected filters), and finally dividing by the sampled air volume.

4.3.2 Outdoor and indoor (household) air temperature

Hourly outdoor temperature and relative humidity data were obtained from the extensive network of meteorological stations in Beijing. We measured indoor temperature in all participant homes prior to blood pressure measurement. In a random 75% subsample of households in each campaign, we also placed a real-time temperature sensor (iButton DS1921G-F5; Thermochron, Maxim Inc., USA) in the room where participants reported spending most of their daytime hours when indoors. Sensors were wall-mounted at a standardized height (~1.5 to 2 meters), away from major heating sources, windows, and doors, and were programmed to log a temperature reading every 125 minutes for up to 4 months to capture the full winter period and early spring weeks when heating may still intermittently occur. Prior to the start of each campaign, we co-located all of the sensors and measured temperature over two days and compared the readings. Sensors recording values >1C from the group median value were excluded from data collection.

4.3.3 Objective measurement of household stove use using sensors

Following methods used in a previous intervention evaluation study in rural China (Clark et al. 2017), we objectively measured household heating stove use in a random sample of households selected, also at random, for either short- or long-term measurement. We measured short-term (24 h) stove use for all household heating stoves in 315 and 227 households in seasons 2 and 3, respectively. Long-term stove use was assessed in 324, 273, and 585 homes in S2, S3, and S4, respectively, for a period of approximately 6 months. We measured stove use using the same real-time temperature data loggers used for indoor temperature (iButton DS1921G-F5; Thermochron, Maxim Inc., USA). Field staff placed the sensors on stoves and programmed them to record surface temperature every 125 minutes, a timing decision based on pilot assessments showing that shorter time intervals did not change the number of heating events detected. Sensors were on the surfaces of biomass and coal-fuelled stoves and radiators. For heat pumps, sensors were placed on the heat exchanger coil on air-to-air units and on the radiator of air-to-water units.

The number and duration of stove combustion events were identified from the temperature data using criteria defined based on the observed changes in the peak shape of the time series temperature curves (i.e., changes in the slope or in absolute temperature compared with the indoor ambient temperature). This approach was specific to heating stoves but developed based on stove use identification for cookstoves in previous studies by us and others (Clark et al. 2017; Ruiz-Mercado et al. 2013; Snider et al. 2018). We developed separate criteria for each stove since heating patterns varied by stove. These criteria were coded into stove-specific algorithms (using R Studio) to systematically identify the number and duration of heating events across households. A random 15% of stove use temperature files were sampled with respect to the stove type and measurement duration (short-term/24 h or long-term/~6 mo), and manually coded to develop the criteria. The number and duration of heating events were identified by the algorithms in the remaining 85% of files. We compared heating periods identified manually with those identified by the algorithm to check for systematic differences and possible overfitting.

4.3.4 Questionnaires

Field staff administered household and individual-level questionnaires to assess household demographic information and educational attainment, household assets, house structure, stove and fuel use patterns (including a complete roster of heating methods and their contributions in each room), and individual health behaviors including exercise frequency, smoking, alcohol consumption, medication use, and clinician-diagnosed health conditions. We used Surveybe computer-assisted personal interview (CAPI) software to collect survey data via handheld electronic tablets. Questions were read to participants in Mandarin-Chinese, and their responses were recorded into tablets.

Prior to the start of data collection, all questions were translated from English into Chinese and then back-translated to English for quality assurance. Many questions were adapted from previous field studies of household energy and blood pressure conducted in rural Beijing or other rural sites in China (Baumgartner et al. 2018; Yan et al. 2020), and all questions were iteratively tested with staff and adapted prior to implementation. Prior to each campaign in this study, the questionnaire and other study measurements were tested in 12 households located in a Beijing village that was eligible for our study but was instead selected for testing. We used the test village to assess whether the questions were understandable and interpreted as intended and to identify any problems with the study measurements or their implementation. Study protocols were subsequently adapted prior to the start of data collection.

In addition to household and individual participant questionnaires, we also conducted village surveys with one representative from each village committee to understand how the policy was implemented in that village and to inquire about any other rural development or health programs being implemented in the village. Committee members answered questions about committee and villager interest in the policy and, for in treated villages, assignment versus application to the policy, any home or village renovations required by the upper-level government prior to heat pump installation, decision-making for the type and brand of heating technology, level of subsidies provided for heaters and electricity, and technical and logistic guidance to villagers.

4.3.5 Blood pressure

Following 5 min of quiet rest, at least three brachial and central systolic (bSBP/cSBP) and diastolic (bDBP/cDBP) blood pressures (BPs) were taken by trained staff at 1 min apart on the participant's supported right arm. We used an automated oscillometric device (BP+; Uscom Ltd, New Zealand) that estimates central pressures from the brachial cuff pressure fluctuations. Central pressures were previously validated against invasive cBP measurements in previous studies (Costello et al. 2015; Lowe et al. 2009). The BP devices were factory calibrated by the manufacturer prior to the start of the first and fourth campaigns. Up to five measurements were taken if the difference between the last two was >5 mmHg or staff were unable to obtain a reading. The BP measurements were conducted in the participant's home and staff were trained to follow strict quality control procedures, including use of an appropriately sized cuff, correct positioning of the arm, both feet on the ground, and ensuring 5 min of quiet rest before measurement. Details are described in the standard operating procedures (SOP): https://osf.io/gmka5. The average of the final two measurements was used for statistical analysis unless only one BP measurement was obtained (n = 13 observations), in which case, a single measurement was used. The time of day, day of the week, and indoor temperature prior to BP measurement were also recorded.

Multiple imputation for covariates in analyses with BP outcomes

Blood pressure was measured at household visits but several key covariates like waist circumference, height, and weight were measured at the clinic visits in S1 and S2. Thus, we were missing covariate information for individuals who were unable to attend the clinic visits (~15-20% of participants in each campaign). Multiple imputation with chained equations (MICE) was conducted to impute missing covariate data for individuals who participated in the household visit but not the clinic visit in analyses with BP outcomes in order to retain observations with BP measurements that would have otherwise been dropped in adjusted models using complete-case analysis. Imputation was performed with the 'MICE' package (van Buuren and Groothuis-Oudshoorn 2011) in R (m = 30 imputation datasets, with 30 iterations each), and the DiD analysis was conducted for each of the 30 datasets. We then used Rubin's Rules to combine point estimates and standard errors while accounting for both within- and between-dataset variances (Rubin 1987).

4.3.6 Self-reported respiratory symptoms and airway inflammation

During questionnaire assessment, participants were asked about chronic airway symptoms including cough, phlegm, wheeze, and tightness in the chest using questions validated for use in Mandarin-Chinese and developed from the standard questionnaires on COPD (Medical Research Council/International Union Against Tuberculosis and Lung Disease) and asthma (International Study of Asthma and Allergies in Childhood). The Mandarin-Chinese questions were extensively piloted with rural and peri-urban Beijing residents to ensure that the health terminology and symptom time patterns were adequate and understandable to the local population.

In a ~25% random subsample of participants, we also measured fractional exhaled nitric oxide (FeNO), a non-invasive and established marker of airway inflammation, using a portable handheld device (Aerocrine, Solna, Sweden) fit with a NIOX VERO® sensor, following ATS recommendations and guidelines (ATS/ERS 2005). Briefly, FeNO measurement was performed with participants in a standing position. They inhaled NO-free air through a mouthpiece with an NO-scrubber attached, followed by controlled expiration for 10 s through the mouthpiece at 50 ± 5 mL/s. A nose clip was used to avoid nasal inhalation, and accurate flow rate was achieved using visual and auditory cues generated by the device. Detailed methods are provided in our previous study of air pollution and FeNO in Beijing adults (Shang et al. 2020). At least two measurements were obtained for each participant.

4.3.7 Blood inflammatory and oxidative stress markers

Trained nurses collected 20 ml of whole blood in a labeled vacutainer via venipuncture using standard techniques (Tuck et al. 2009). Details are described in our published SOP. Briefly, fasting blood samples were collected by experienced phlebotomists (nurses) in the morning and stored at 4-10°C prior to centrifugation. Two serum aliquots from each participant were then placed in a -30°C freezer for temporary storage. Collection-to-storage time was <4 hrs for all samples in both campaigns where blood samples were collected. Within 3-5 days of collection, the samples were transported in styrofoam containers with dry ice to a -80°C freezer with a backup generator and alarm system at Peking University.

The first aliquot was analyzed for glucose and a complete lipid profile within two months of collection, and results were communicated to participants. The second aliquot was stored in the -80°C freezer for analysis of biomarkers of systemic inflammation [C-reactive protein (CRP), interleukin-6 (IL-6), tumour necrosis factor alpha (TNF-α) and oxidative stress [8-hydroxy-2'-deoxyguanosine (8-OHdG) and malondialdehyde (MDA)] at the University of the Chinese Academy of Sciences between July and September of 2023. These biomarkers were selected because they are associated with the development of cardiovascular disease and events (e.g., Danesh 2008; Pearson 2003; Ridker 2000; 2001; ERF 2012), and both acute and longer-term exposures to air pollution have been associated with changes in inflammatory and oxidative stress markers (e.g., Pope 2004; Rückerl 2007; Rich 2012; Kipen 2010; Huang 2012).

We followed standard methods for analysis (FDA Guidance, 2018). For inflammatory markers (IL-6, TNF- , CRP), the optic densities (OD) of all samples were measured using an automated ELISA reader. Every plate had 8 standard samples used to generate a standard curve that related OD and standard inflammatory marker concentration. A standard curve for each microplate was generated by a computer software program based on a 4-parameter method. Each plate included at least 3 control samples to ensure the stability of standard curves. All samples, standards, and controls were measured in duplicate, and the average was used for statistical analysis. For oxidative stress biomarkers (MDA and 8-OHdG), the chromatographic peak areas of all samples were measured using HPLC with UV detector and HPLC-MS/MS. Every plate had 7 standard samples used to generate a standard curve that related peak area and concentration of each standard oxidative stress marker. A standard curve for each plate was generated using a computer software program based on a linear method. Each plate included at least 3 control samples to ensure the stability of standard curves. Standards and controls were measured in duplicate and samples were measured once due to high precision in a pilot study (Food and Drug Administration 2018).

4.3.8 Anthropometric measurements.

Body weight, height, and waist circumference were measured at the clinic visit in the first two campaigns and in participant homes in the last campaign. Weight was measured in light indoor clothing without shoes in kilograms to one decimal place, using standing scales supported on a steady surface. The scales were calibrated prior to the start of each campaign, and the same staff member stepped on the scale each morning to ensure that it was functioning properly. Height was measured without shoes in centimeters to one decimal place with a stadiometer. Waist circumference was measured without clothing obstruction at one centimeter above the participant's navel at minimal respiration in centimeters to one decimal place. The measuring tape was replaced at the start of each campaign to avoid stretching.

4.4 Measuring policy impacts

To understand how Beijing's policy works we used a difference-in-differences (DiD) design (Callaway 2020), leveraging the staggered rollout of the policy across multiple villages to estimate its impact on health outcomes and understand the mechanisms through which it works. Simple comparisons of treated and untreated (i.e., control) villages after the CBHP policy has been implemented are likely to be biased by unmeasured village-level characteristics (e.g., migration, average winter temperature) that are associated with health outcomes. Similarly, comparisons of only treated villages before and after exposure to the program are susceptible to bias by other factors associated with changes in outcomes over time (i.e., secular trends, impacts of the COVID-19 pandemic). By comparing the *changes* in outcomes among treated villages to the *changes* in outcomes among untreated villages, we can control for any unmeasured time-invariant characteristics of villages as well as any general secular trends affecting all villages that are unrelated to the policy.

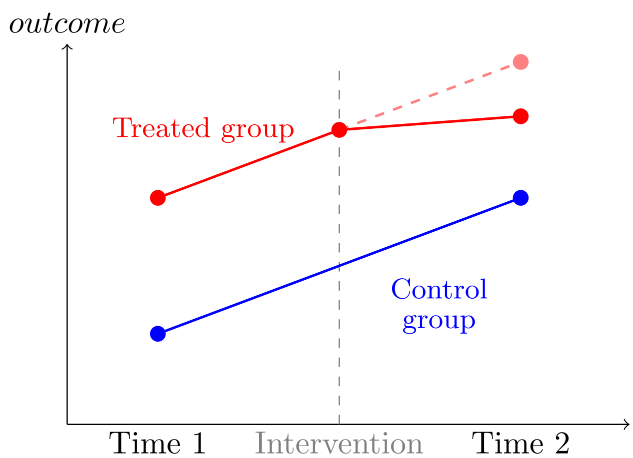


Figure 3: Stylized example of difference-in-differences

The DiD design compares outcomes before and after an intervention in a treated group relative to the same outcomes measured in a control group. The control group trend provides the crucial "counterfactual" estimate of what would have happened in the treated group had it not been treated. By comparing each group to itself, this approach helps to control for both measured and unmeasured fixed differences between the treated and control groups. By measuring changes over time in outcomes in the control group unaffected by the treatment, this approach also controls for any unmeasured factors affecting outcome trends in both treated and control groups. This is important since there are often many potential factors affecting outcome trends that cannot be disentangled from the policy if one only studies the treated group (as in a traditional pre-post design).

The canonical DiD design (Card and Krueger 1994) compares two groups (treated and control) at two different time periods (pre- and post-intervention, Figure 3). In the first time period both groups are untreated, and in the second time period one group is exposed to the intervention. If we assume that the differences between the groups would have remained constant in the absence of the intervention (parallel trends assumption), then an unbiased estimate of the impact of the intervention in the post period can be calculated by subtracting the pre-post difference in the untreated group from the pre-post difference in the treated group.

However, when multiple groups are treated at different time periods, the most common approach has been to use a two-way fixed effects model to estimate the impact of the intervention which controls for secular trends and differences between districts. However, recent evidence suggests that the traditional two-way fixed effects estimation of the treatment effect may be biased in the context of heterogeneous treatment effects (Callaway and Sant'Anna 2021; Goodman-Bacon 2021).

4.5 Measuring pathways and mechanisms

To estimate how much of the CBHP intervention may work through different mechanisms, we used causal mediation analysis. Causal approaches to mediation attempt to discern between, and clarify the necessary assumptions for identifying, different kinds of mediated effects. Taking as an example the DAG in Figure 4, with T as the policy, M as $\mathrm{PM}_{2.5}$, and Y as systolic blood pressure, we can define the controlled direct effect (CDE) as the effect of the CBHP policy on systolic blood pressure if we fix the value of $\mathrm{PM}_{2.5}$ to a certain reference level for the entire population. For example, we can estimate the impact of the policy on health outcomes while holding $\mathrm{PM}_{2.5}$ at a uniform level of average background exposure, or some other hypothetical level.

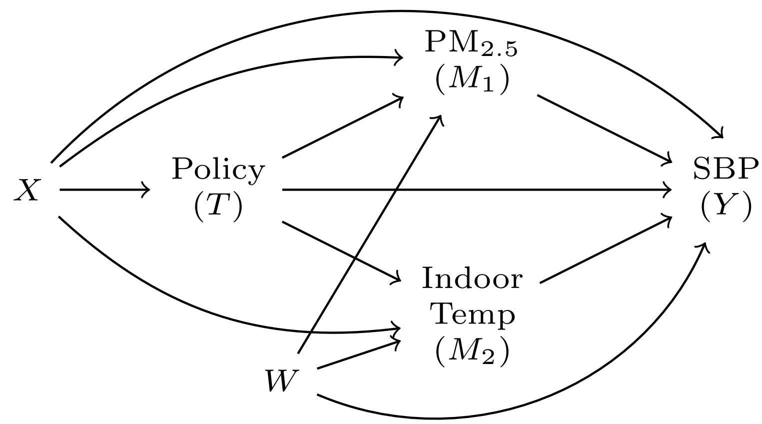


Figure 4: Hypothetical Directed Acyclic Graph showing direct and indirect effects with outcome (Y), pre-treatment covariates (X), policy (T), multiple mediators (M_1, M_2), as well as covariates for the mediators (W).

Although other mediated effects such as "natural" direct and indirect effects are theoretically estimable (VanderWeele 2015), they involve challenging "cross-world" assumptions that are difficult to anchor in policy (Naimi et al. 2014). Other approaches to mechanisms have focused on principal stratification (e.g., Zigler et al. 2016), although conceptual difficulties with identifying the (unverifiable) principal strata make it challenging for questions of mediation. Because controlled direct effects are considered more directly policy relevant for public health, we focus on estimating these mediated quantities.

5 Data Analysis

To understand how the policy's impact on health may be mediated by different potential mediators, we need to estimate first the total effect of the policy on the outcomes, as well as the CDEs with adjustment for potential mediators. As discussed above, in order for the mediators to 'explain' the total effects of the policy on health, the policy should affect the mediators, and the mediators should also affect the outcomes.

5.1 Total Effect

To estimate the total effect of the policy we used a DiD analysis that accommodates staggered treatment rollout. To allow for heterogeneity in the context of staggered rollout we used 'extended' two-way fixed effects (ETWFE) models (Wooldridge 2021) to estimate the total effect of the CBHP policy. The mean outcome (replaced by a suitable link function $g(\cdot)$ for binary or count outcomes) was defined using a set of linear predictors:

$$Y_{ijt} = g(\mu_{ijt}) = \alpha + \sum_{r=q}^{T} \beta_r d_r + \sum_{s=r}^{T} \gamma_s fs_t + \sum_{r=q}^{T} \sum_{s=r}^{T} \tau_{rt}(d_r \times fs_t) + \varepsilon_{ijt} \tag{1}$$

where Y_{ijt} is the outcome for individual i in village j at time t, d_r represent treatment cohort dummies, i.e., fixed effects for cohorts of villages that were first exposed to the policy at the same time q (e.g., in 2019, 2020, or 2021), fs_t are time fixed effects corresponding to different winter data collection campaigns (2018-19, 2019-20, or 2021-22), and τ_{rt} are the cohort-time $ATTs$. The ETWFE and other approaches that allow for several (potentially heterogenous) treatment effects may also be averaged to provide a weighted ATT. Several potential possibilities are feasible, including weighting by treatment cohorts or time since policy adoption (Goin and Riddell 2023).

5.2 Mediation Analysis

As noted above, with respect to the mediation analysis we are chiefly interested in the CDE, which can be derived by adding relevant mediators M to this model. If we also allow for exposure-mediator interaction and potentially allow for adjustment for confounders W of the mediator-outcome effect, we can extend equation Equation 1 as follows:

$$
\begin{aligned}
Y_{ijt} = g(\mu_{ijt}) = \alpha + \sum_{r=q}^{T} \beta_r d_r + \sum_{s=r}^{T} \gamma_s fs_t + \sum_{r=q}^{T} \sum_{s=r}^{T} \tau_{rt}(d_r \times fs_t) \\
+ \delta M_{it} + \sum_{r=q}^{T} \sum_{s=r}^{T} \eta_{rt}(d_r \times fs_t \times M_{it}) + \zeta \mathbf{W} + \varepsilon_{ijt}
\end{aligned} \tag{2}
$$

where now δ is the conditional effect of the mediator M at the reference level of the treatment (again, represented via the series of group-time interaction terms), and the collection of η terms are coefficients for the product terms allowing for mediator-treatment interaction. Finally, ζ is a vector of coefficients for the set of confounders contained within $\mathbf{W}$.

As noted above, in the staggered DiD framework that allows for heterogeneity, we do not have a single treatment effect but a collection of group-time treatment effects that may be averaged in different ways. This extends to the estimation of the CDE, in which case we will also have several CDEs that can be averaged to make inferences about the extent to which the policy's impact is mediated by $PM_{2.5}$. Based on the setup in Equation 2 the CDE is estimated as: $\delta + \eta_{rt} MT$. In the absence of interaction between the exposure and the mediator (i.e., $\eta_{rt} = 0$) the CDE will simply be the estimated treatment effects $\sum_{r=q}^{T} \sum_{s=r}^{T} \tau_{rt}$, i.e., the effect of the policy holding M constant. For a valid estimate of the CDE we must account for confounding of the mediator-outcome effect, represented by W in the equation above. The inclusion of baseline measures of both the outcome and the proposed mediators inherent in our DiD strategy help to reduce the potential for unmeasured confounding of the mediator-outcome effect (Keele et al. 2015). Given the large number of outcomes of interest in this study, as well as the potential for heterogeneous treatment effects, we limited the mediation analysis to health outcomes for which we observed a total effect of the CBHP policy.

5.3 Identification of potential confounders and model covariates

DiD is a strong analytical approach that already minimizes the risk of confounding, where cohort-fixed effects control for measured and unmeasured time-constant factors that may differ between treatment cohorts (e.g., genetics, altitude), and time-fixed effects control for secular trends that affect all treatment cohorts similarly over the study period (e.g., background improvements in ambient air quality or household transition to more efficient heating).

For models estimating the effect of the policy on health outcomes, we used directed acyclic graphs (DAGs) to identify potential time-varying causes of both treatment by the policy and our study outcome(s) that could differ between treatment groups, and adjusted for those potential confounders in the regression models. For the mediation analysis, we identified potential mediator-outcome confounders using the same approach. These variables were identified from the relevant peer-reviewed literature and our team's substantive knowledge about the CBHP policy. In the multivariable models, we also adjusted for strong predictors of the outcome that were not affected by treatment, and thus not confounders, to improve model precision. The covariates included in each of the models for cardiovascular and respiratory health outcomes are provided in the tables.

For air pollution outcomes, we considered the following covariates: village population and total number of households in the village; temperature, relative humidity, wind direction, wind speed, boundary layer height; home area and home area heated; home insulation; smoking status of participant and whether or not they lived with a smoker; whether or not the household reported using wood (i.e., biomass) for household energy activities, and if so, self-reported quantity of wood. Potential non-linearity between continuous covariates and our study outcomes were evaluated using cubic splines. Ultimately, the following covariates were included in the final DiD models for outdoor, indoor, and personal exposures to air pollution, based on whether measurable changes in the covariate over time were observed. For the final adjusted DiD model for personal exposure 'mixed combustion' source contributions, the following covariates were included: temperature (represented by a spline with 2 degrees of freedom); participant smoking status; and whether or not the household reported using biomass fuel. For the final adjusted DiD model for outdoor (community) 'mixed combustion' source contributions, the following covariates were included: total number of households in the village; village population; and ambient relative humidity (represented by a spline with 2 degrees of freedom).

6 Results

We retained all 50 study villages during this four-year longitudinal assessment of village treatment by the CBHP policy, though we were only able to visit 41 villages in winter 2020-21 (S3) and were limited to village and household-level measurements of air quality, indoor temperature, and stove use in that campaign due to travel restrictions during the COVID pandemic.

By S2, S3, and S4 there were a cumulative total of 10, 17, and 20 (out of 50 total) study villages treated by the CBHP policy, respectively. All of the treated villages in our study selected to install electric-powered air-source heat pumps with 200 RMB per meter square (up to 24,000 RMB) in subsidies and were also provided with 80% night-time electricity subsidies up to 10,000kWh per heating season. To limit coal use, villages enrolled in the policy were no longer allowed to place orders for subsidized coal with the district-level governments that manage the procurement and distribution of coal for residential heating in Beijing. In addition, village committee leaders in treated villages reported feeling accountable to the Environmental Protection Department for limited coal-related air pollution, and were motivated to encourage residents to not burn coal. Some villages were equipped with government air pollution monitors and the Environmental Protection Department conducted village inspections and issued warnings about coal burning. Households burning coal in treated villages were at risk of losing their electricity subsidy.

Table 2: Demographic and health characteristics of participants in each study campaign.

Characteristic	**Wave 1 (2018-19) N=1003**	**Wave 2 (2019-20) N=1110**	**Wave 4 (2021-22) N=1028**
Female, n (%)	580 (57.8)	653 (58.8)	612 (59.5)
Current smoker, n (%)	257 (25.6)	295 (26.6)	265 (25.8)
Any smoke exposure, n (%)	788 (78.6)	897 (80.8)	843 (82)
Age in years, Mean (SD)	60.7 (9.2)	61.4 (9.1)	63.1 (9)
BMI in kg/m^2, Mean (SD)	26.1 (3.7)	25.7 (3.5)	26.1 (4)
Waist circumference in cm, Mean (SD)	86.8 (10.2)	87.4 (9.4)	91.4 (10.7)

Table 3: Demographic and health characteristics of participants who contributed to different numbers of campaigns.

Characteristic	**1 Wave N=365**	**2 Waves N=443**	**3 Waves N=630**
Female, n (%)	211 (57.8)	253 (57.1)	370 (58.7)
Current smoker, n (%)	110 (30.1)	117 (26.4)	161 (25.6)
Any smoke exposure, n (%)	288 (78.9)	360 (81.3)	498 (79)
Age in years, Mean (SD)	59.9 (9.2)	60.5 (8.8)	61.3 (9.1)
BMI in kg/m^2, Mean (SD)	26.3 (3.6)	25.8 (3.5)	26.1 (3.7)
Waist circumference in cm, Mean (SD)	90.3 (9.8)	86.5 (10)	86.9 (10.4)

Appendix Figure A3 shows the participation of villages, households, and participants across the four waves of data collection. We conducted measurements in over 1000 participants in each of the three measurement campaigns that included individual-level measurements. In total, we enrolled 1432 participants into the study, of which 630 (43%) participated in all three campaigns, 443 (31%) participated in two campaigns and 365 (25%) participated in a single campaign. We did not observe any notable differences in demographic characteristics or health behaviors between participants who contributed to a different number of campaigns (Table 2) or between participants in each of the three campaigns with individual measurements (Table 3).

6.1 Description of study sample

Table 4 shows the distribution of selected demographic, health, and environmental characteristics from the baseline survey, prior to any villages being enrolled in the CBHP policy. We provide means and standard deviations separately for villages that eventually enter into the policy with those that

Table 4: Descriptive statistics for selected demographic, health, and environmental measures at baseline, by treatment status

	Never treated (N=603)		Ever treated (N=400)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	Std. Error
Demographics:						
Age (years)	59.9	9.4	60.4	9.2	0.5	0.6
Female (%)	59.5	49.1	59.1	49.2	-0.4	3.2
No education (%)	11.5	31.9	12.3	32.9	0.9	2.1
Primary education (%)	75.5	43.0	77.6	41.7	2.1	2.8
Secondary+ education (%)	12.6	33.2	9.8	29.7	-2.9	2.0
Health measures:						
Never smoker (%)	61.9	48.6	59.5	49.1	-2.4	3.2
Former smoker (%)	11.9	32.4	15.1	35.8	3.2	2.2
Current smoker (%)	26.2	44.0	25.4	43.6	-0.8	2.8
Never drinker (%)	55.9	49.7	52.5	50.0	-3.4	3.2
Occasional drinker (%)	26.0	43.9	25.5	43.6	-0.5	2.8
Daily drinker (%)	17.8	38.3	21.9	41.4	4.1	2.6
Systolic (mmHg)	131.4	16.8	128.7	14.3	-2.7	1.0
Diastolic (mmHg)	82.7	11.6	82.1	11.3	-0.6	0.8
Waist circumference (cm)	87.7	10.5	85.4	9.5	-2.3	0.8
Body mass index (kg/m2)	26.3	3.7	25.8	3.6	-0.5	0.3
Frequency of coughing (%)	18.7	39.0	19.7	39.8	1.0	2.6
Frequency of wheezing (%)	6.2	24.2	6.6	24.8	0.3	1.6
Shortness of breath (%)	29.2	45.5	34.3	47.5	5.1	3.0
Chest trouble (%)	11.6	32.0	14.1	34.9	2.5	2.2
Any respiratory problem (%)	50.6	50.0	54.3	49.9	3.7	3.2
Environmental measures:						
Temperature (°C)	13.8	3.6	13.5	3.3	-0.3	0.2
Personal PM2.5 (ug/m3)	150.2	300.3	103.8	107.3	-46.3	19.1

never do so. As noted above, although our DiD identification strategy allows for fixed differences between treated and untreated villages, overall the differences at baseline are generally small and the groups seem well balanced on most measures, with the exception of personal exposure to $PM_{2.5}$, which was lower in villages that were eventually treated.

6.2 Summary of PM and BC measurements

At baseline, fine particulate matter ($PM_{2.5}$) and black carbon (BC) concentrations were higher, on average, for personal exposures compared with outdoor concentrations. From Season 2 onward, with the inclusion of indoor air pollution measurements, personal exposure air pollution concentrations were still higher than indoor or outdoor concentrations, with indoor levels being higher than outdoors (Table 5). This trend (personal > indoor > outdoor) was observed among households in treated and untreated villages. Personal, indoor, and outdoor geometric mean (95% confidence interval) concentrations of $PM_{2.5}$ were 72 (65, 80), 45 (39, 53), and 31 (28, 35)[a], respectively, and elevated relative to health-based guidelines. The current World Health Organization (WHO) guidelines state that annual average concentrations of $PM_{2.5}$ should not exceed 5 µg/m^3, while 24-hour average exposures should not exceed 15 µg/m^3 for more than 3 to 4 days per year (Organization 2021). Interim targets have been set to support the planning of incremental milestones toward cleaner air, particularly for cities, regions, and countries with higher air pollution levels. For $PM_{2.5}$, the four interim (IT) targets for annual and 24-h means are: IT-1: 35 and 75 µg/m^3; IT-2: 25 and 50 µg/m^3; IT-3: 15 and 37.5 µg/m^3; and IT-4: 10 and 25 µg/m^3 (Organization 2021). In our study, baseline personal exposures to $PM_{2.5}$ fell between IT-1 and IT-2, indicating considerable opportunity for air quality improvement from intervention.

We also present the geometric and arithmetic means (and 95% confidence intervals) for $PM_{2.5}$ and BC in each seasonal measurement campaign (Table 5). Season 3 (2020/2021) was a partial campaign that took place over a time period impacted by the COVID-19 pandemic and did not involve filter-based air pollution sample collection.

6.3 Policy uptake

Each year of the study, participants reported the types of fuels and stoves and the amount of fuel used for space heating in winter. Based on these data, heating energy types were classified into four categories: exclusive use of a heat pump ('Heat pump exclusively'), use of a heat pump and a biomass-fueled kang ('Heat pump with kang'), use of solid fuel heater with an electric heating devices other than heat pumps ('Solid fuel with electric heating appliances'), and exclusive use of solid fuel ('Solid fuel stove exclusively'). In villages treated by the policy, Figure 5 shows meaningful transitions from solid fuel to electric-powered heat pumps for all treatment cohorts. For example, the proportion of households in the group treated in 2019 (S2) using heat pumps increased from 3% in S1 to 93% in S2 and 96% in S4. Conversely, use of coal stoves decreased from 97% in S1 to

Table 5: Arithmetic and geometric means for air pollutant concentrations (micrograms per cubic meter) by season.

			Season 1		Season 2		Season 3		Season 4	
			Est.	CI	Est.	CI	Est.	CI	Est.	CI
Personal										
Filter-derived	24h PM2.5	Mean	117	[105, 129]	97	[87, 107]			84	[72, 97]
		GM	72	[65, 80]	59	[53, 65]			47	[42, 52]
	24h BC	Mean	4	[3.5, 4.4]	3.5	[2.7, 4.2]			3.7	[2.9, 4.5]
		GM	2.6	[2.4, 2.8]	1.9	[1.7, 2.1]			1.7	[1.5, 1.9]
Indoor										
Sensor-derived	Seasonal PM2.5	Mean			94	[84, 104]	84	[75, 94]	67	[60, 75]
		GM			71	[65, 78]	63	[57, 70]	47	[42, 52]
Filter-derived	24h PM2.5	Mean			69	[59, 79]			59	[49, 69]
		GM			45	[39, 53]			33	[27, 40]
	24h BC	Mean			2.3	[1.8, 2.8]			2.8	[2.1, 3.4]
		GM			1.6	[1.3, 2.0]			1.6	[1.3, 1.9]
Outdoor										
Sensor-derived	Seasonal PM2.5	Mean	47	[45, 48]	55	[54, 56]	23	[22, 23]	33	[32, 34]
		GM	36	[35, 37]	40	[39, 41]	33	[32, 34]	22	[22, 23]
Filter-derived		Mean	38	[34, 42]	38	[34, 41]			26	[24, 28]
		GM	33	[29, 36]	30	[28, 32]			22	[21, 24]
	Seasonal BC	Mean	1.5	[1.3, 1.6]	1.4	[1.3, 1.5]			1.2	[1.1, 1.2]
		GM	1.3	[1.1, 1.4]	1.1	[1.0, 1.2]			1	[0.9, 1.1]

Note: Est. = Estimate, CI = 95% CI, GM = Geometric Mean

8% in S2 and 3% in S4. We observed similar stove use transitions for households in villages treated in 2020 (S3). In the three villages treated in 2021, we observed overall less exclusive use of the heat pump and a slightly larger proportion of households continuing to use coal.

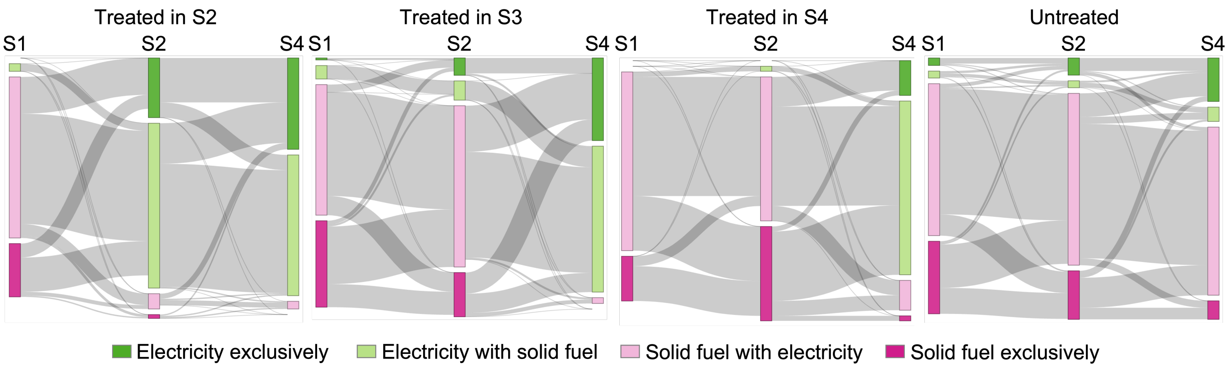


Figure 5: Transitions to different energy sources across study seasons

We also observed a substantial decline in the amount of self-reported coal used in villages treated by CBHP policy (Appendix Figure A4), though the reduction in coal use was smaller with each subsequent treatment cohort (Appendix Table A1). Biomass (i.e., wood logs/twigs or charcoal), usually burned in kangs for both cooking and space heating, was not expressly targeted by the CBHP policy. We observed declines in self-reported biomass use in villages treated in 2019 and 2020 but there was a small increase in biomass consumption in the cohort treated last (2021).

In never treated villages, we also observed a transition from solid fuel to clean energy over the four year study but it was much slower than in treated villages. The proportion of households that reported using electric heat pumps increased from 5% in S1 to 10% in S2 and 25% in S4, and those who adopted heat pumps tended to use them exclusively. Commensurately, the reported expenditures on electricity increased gradually over time in the untreated villages (Appendix Figure A4). The percentage of untreated households using solid fuel with other types of electric devices remained relatively stable, ranging from 64% to 70% across campaigns. Self-reported use of biomass also remained stable, at approximately one ton of fuel each winter, whereas exclusive use of solid fuel decreased from 30% in S1 to 7% in S4.

6.4 Aim 1: Policy impacts and potential mediation

6.4.1 Impact of policy on potential mediators

In estimating the treatment effect on indoor and outdoor air pollution, we evaluated both 24-h mean values (specifically, the same 24-h period when personal exposure samples were collected in each village) and seasonal mean values (with 'season' defined from Jan. 15th to Mar. 15th) of $PM_{2.5}$ data collected in each village. For estimating the treatment effect on personal exposure to $PM_{2.5}$ and black carbon (BC), the results from the filter-based measurements that were collected for a 24-h period were used for analysis. We estimated the basic ETWFE models for outdoor, personal, and indoor measures of air pollution. ETWFE models were further adjusted for covariates, including temperature, relative humidity, wind speed, boundary layer height, wind direction, and the mean quantity of wood burned in each village (for outdoor measures of air pollution); outdoor temperature, dewpoint, household smoking status, and the number of residents in each household (for indoor measures of air pollution); and outdoor temperature, dewpoint, household smoking status, and the number of residents in each household (for personal measures of air pollution).

Treatment was associated with similar reductions in both seasonal and 24-h indoor $PM_{2.5}$ means (Table 6). The average marginal effect (ATT) from the basic ETWFE model shows that exposure to the CBHP policy reduced 24h indoor $PM_{2.5}$ by -19 µg/m^3 (95%CI: -23, 61). After adjusting outdoor temperature, dewpoint, household smoking status, and the number of residents in each household, the ATT decreased to -14 µg/m^3 (95%CI: -54, 26). The impact was stronger on seasonal indoor $PM_{2.5}$, with an average ATT of -36 µg/m^3 (95%CI: -61, -12) that was robust to covariate adjustment. This finding likely reflects the direct benefit of the policy in replacing coal stoves and air quality improvement.

Overall we found little evidence of an impact of the CBHP policy on 24-h and seasonal outdoor (local community-level) $PM_{2.5}$ or personal exposures to $PM_{2.5}$ and BC. Treatment was associated with lower, but statistically imprecise, personal 24-h BC exposures. This finding would be consistent with the expectation that the policy contributed to reducing air pollutant emissions from solid fuel burning, as BC serves as a potential indicator of such combustion, particularly in our rural and peri-urban study villages.

6.4.2 Impact of policy on health outcomes

Table 7 shows the impacts of the policy on blood pressure in basic ETWFE models and models further adjusted for age, sex, waist circumference, smoking, alcohol consumption, and use of blood pressure medication. Overall exposure to the CBHP policy demonstrated reductions in blood pressure of approximately 1.5 mmHg for both systolic and diastolic BP, but we found little evidence of a meaningful impact on pulse pressure or BP amplification. The effects on brachial and central blood pressures were similar.

Table 7 shows the impacts on self-reported chronic respiratory symptoms categorized as any symptoms and separately for each individual symptom type. In both basic and covariate-adjusted ETWFE models, exposure to the CBHP policy reduced self-report of any poor respiratory symptoms by around 7 percentage points. This was largely through reductions in reports of having chest trouble or difficulty breathing on several or most days of the week.

Table 6: Treatment effect on outdoor and indoor $PM_{2.5}$, personal exposure to $PM_{2.5}$ and black carbon, and measures of indoor temperature. Outdoor and indoor $PM_{2.5}$ were derived from sensor measurements after being adjusted based on co-located gravimetric $PM_{2.5}$ measurements. 24h indicates the mean $PM_{2.5}$ concentrations during the 24 hours when personal exposure samples were collected in each village. 'Seasonal' indicates the seasonal mean $PM_{2.5}$ concentrations in each village, from Jan. 15th to Mar. 15th.

		DiD		Adjusted DiD[a]	
		ATT	(95% CI)	ATT	(95% CI)
Air pollution (μg/m^3)					
Personal	PM2.5	-2.09	(-29.38, 25.2)	1.95	(-23.34, 27.23)
	Black carbon	-0.46	(-1.73, 0.81)	-0.43	(-1.67, 0.81)
Indoor	Daily	-19.10	(-60.56, 22.35)	-14.20	(-53.94, 25.54)
	Seasonal	-35.11	(-59.36, -10.85)	-36.19	(-60.74, -11.65)
Outdoor	Daily	-0.11	(-5.86, 5.64)	-1.73	(-9.26, 5.81)
	Seasonal	3.14	(-3.1, 9.38)	0.36	(-6.27, 6.99)
Indoor temperature (°C)					
Point	Mean	1.96	(0.96, 2.96)	1.96	(0.96, 2.96)
Seasonal	Mean (all)	0.64	(0, 1.29)	0.64	(0, 1.29)
	Mean (daytime)	0.82	(-0.08, 1.72)	0.82	(-0.08, 1.72)
	Mean (heating season)	1.80	(0.96, 2.64)	1.80	(0.96, 2.64)
	Mean (daytime heating season)	1.85	(0.97, 2.73)	1.85	(0.97, 2.73)
	Min. (all)	3.83	(2.26, 5.39)	3.83	(2.26, 5.39)
	Min. (heating season)	3.72	(2.19, 5.25)	3.72	(2.19, 5.25)

Note: ATT = Average Treatment Effect on the Treated, DiD = Difference-in-Differences, ETWFE = Extended Two-Way Fixed Effects.

[a] ETWFE models for air pollution outcomes were adjusted for household size, smoking, outdoor temperature, and outdoor humidity. Temperature models not additionally adjusted.

Table 7: Overall impacts of the 'coal-to-clean energy' policy on blood pressure, respiratory outcomes, and inflammatory markers

		DiD		Adjusted DiD[a]	
		ATT	(95% CI)	ATT	(95% CI)
Blood pressure (mmHg)					
Systolic BP	Brachial	-0.79	(-2.63, 1.04)	-1.40	(-3.31, 0.51)
	Central	-1.04	(-2.82, 0.73)	-1.56	(-3.40, 0.28)
Diastolic BP	Brachial	-1.29	(-2.62, 0.04)	-1.60	(-2.96, -0.25)
	Central	-1.35	(-2.66, 0.04)	-1.66	(-2.97, -0.34)
Pulse Pressure	Brachial	0.50	(-0.71, 1.70)	0.21	(-1.00, 1.41)
	Central	0.31	(-0.85, 1.46)	0.10	(-1.01, 1.20)
BP Amplification x10	Pulse pressure	0.10	(-0.12, 1.40)	0.00	(-1.20, 1.20)
	Systolic BP	0.20	(-0.20, 0.50)	0.10	(-0.20, 0.40)
Respiratory outcomes					
Self-reported (pp)	Any symptom	-7.38	(-13.98, -0.77)	-7.86	(-14.63, -1.09)
	Coughing	-1.59	(-6.41, 3.23)	-1.98	(-6.8, 2.84)
	Phlegm	-1.22	(-5.58, 3.15)	-1.82	(-6.34, 2.69)
	Wheezing attacks	-0.22	(-3.97, 3.52)	-0.14	(-3.85, 3.57)
	Trouble breathing	-4.98	(-11.81, 1.84)	-4.62	(-11.59, 2.35)
	Chest trouble	-6.63	(-12.51, -0.76)	-6.36	(-12.14, -0.59)
Measured	FeNO (ppb)	0.17	(-2.24, 2.58)	0.55	(-2.13, 3.13)
Inflammatory markers (%)					
	IL6	6.80	(-12.2, 30.0)	5.90	(-13.8, 30.2)
	TNF-alpha	24.30	(-1.3, 56.4)	24.70	(-0.9, 54.2)
	CRP	2.70	(-19.8, 31.6)	3.80	(-19.4, 33.6)
	MDA	7.60	(-8.7, 26.9)	6.50	(-9.7, 25.5)

Note: ATT = Average Treatment Effect on the Treated, DiD = Difference-in-Differences, pp = percentage points, ppb = parts per billion.

[a] Blood pressure models adjusted for age, sex, waist circumference, smoking, alcohol consumption, and use of blood pressure medication.

Table 7 also shows the impacts of the CBHP on measured airway inflammation (FeNO), which was conducted in a sub-sample of 511 participants, including 274 participants with one measurement, 142 with two measurements, 95 participants with 3 measurements. We did not find evidence that exposure to the policy affected changes in FeNO in the covariate-adjusted ETWFE model (0.5 ppb, 95%CI: -2.1, 3.1). There was some evidence of heterogeneity in the FeNO effects of the policy by treatment cohort, though the confidence intervals for each of the cohort-specific effects were large and overlapping. Our results did not change with sensitivity analyses that included a log-transformed FeNO outcome and limiting the analysis to participants with at least two repeated measurements and to those who participated in all three campaigns (SI Table X)

6.4.3 Mediated impact on blood pressure

As noted above, we aimed to assess whether any health impacts of the CBHP policy may work specifically through pathways involving changes in $PM_{2.5}$ and indoor temperature. Below we show results from several mediation models. We evaluated potential mediation for each mediator (indoor temperature and personal exposure to $PM_{2.5}$) separately and in a single model accounting for multiple mediators, and we set the values of both mediators to the WHO mean annual interim $PM_{2.5}$ and indoor temperature guidelines. For mediation analysis, we focused on BP outcomes for which we observed an effect of the policy. In Table 8 we show that conditioning on indoor PM and indoor temperature largely explains the entire total effect of the CBHP policy on blood pressure for systolic BP, and roughly half of the total effect for diastolic BP.

Table 8: Controlled direct effects for the CBHP policy

	Adjusted Total Effect[a]		CDE Mediated By:[b]					
			Indoor PM		Indoor Temp		PM + Temp	
	ATT	(95%CI)	ATT	(95%CI)	ATT	(95%CI)	ATT	(95%CI)
Brachial SBP	-1.40	(-3.31, 0.51)	-1.05	(-3.12, 1.02)	-0.46	(-2.29, 1.36)	-0.03	(-2.04, 1.97)
Central SBP	-1.56	(-3.40, 0.28)	-1.15	(-3.20, 0.89)	-0.68	(-2.36, 1.00)	-0.20	(-2.11, 1.70)
Brachial DBP	-1.60	(-2.96, -0.25)	-1.40	(-2.97, 0.16)	-1.14	(-2.33, 0.06)	-0.88	(-2.30, 0.54)
Central DBP	-1.66	(-2.97, -0.34)	-1.40	(-2.96, 0.16)	-1.32	(-2.50, -0.14)	-1.02	(-2.45, 0.41)

Note: Results combined across 30 multiply-imputed datasets. ATT = Average Treatment Effect on the Treated, CDE = Controlled Direct Effect, DBP = Diastolic blood pressure, SBP = Systolic blood pressure.
[a] Adjusted for age, sex, waist circumference, smoking, alcohol consumption, and use of blood pressure medication.
[b] Mediators were set to the mean value for untreated participants at baseline.

6.5 Aim 2: Source contributions

Source analysis for this study was conducted using data from all eligible outdoor PM and personal PM samples. Eligible samples were those for which $PM_{2.5}$ mass and chemical components were quantified. We evaluated factors contributing to community-outdoor and personal exposure $PM_{2.5}$ using the U.S. EPA's source apportionment model PMF (positive matrix factorization) 5.0, which has been widely used for similar analyses in China (Gao et al. 2018; Liu et al. 2017; Tao et al. 2017). As an optimum PMF result depends on the appropriate number of input factors, sensitivity analysis using a range of factors (e.g., range of 3 to 7 factors, based on a combination of the species that we have and our field-based observations and sources that have been identified previously in our study region) were conducted to examine the impact of a different number of factors on the model results. Detailed information on the procedures of PMF analysis can be found elsewhere (Wang et al. 2016; Zíková et al. 2016). Briefly, the scree plot from our principal component analysis indicated that solutions of between 3 and 5 factors (+/- 1) would be most appropriate, further supporting our evaluation of 3 to 6 factor solutions from PMF. As there was no indication that even moving from five factors to six factors would improve our solution; therefore, a seven factor solution would not make sense to investigate further (Figure 6).

The chemical analysis data used as the input for the PMF model were dispersion normalized prior to inclusion in the model. PMF works by using covariance of compositional variables to separate sources of ambient PM. However, atmospheric dilution also induces covariance. Dilution can be quantified in terms of a ventilation coefficient (VC) and used to normalize the input chemical concentrations and uncertainties in the original data matrix on a sample by sample basis. The dispersion normalized concentrations and uncertainties are used as the input to PMF analysis. Dispersion normalization, as conducted in this study, is a relatively new application of this conceptual framework (Dai et al. 2020), developed to adjust for wind speed (dispersion in the x-y plane) and boundary layer height (dispersion in the z-axis). This process involves first calculating the sample specific ventilation coefficient by multiplying the average wind speed by the average boundary layer height over the sampling duration. The average ventilation coefficient is also calculated for the village by averaging all the ventilation coefficients. The dispersion normalized concentration for any species in any sample is equal to the species concentration in that sample multiplied by the ventilation coefficient for that sample and divided by the average ventilation coefficient for that village. Dividing by the average ventilation coefficient for that village helps curtail any extreme concentrations driven by an outlier in the sample ventilation coefficient.

The meteorological data included hourly boundary layer height, 2-m temperature, 2-m dew point temperature, and 2-m horizontal wind speed components (u, v), which were obtained from the European Center for Midrange Weather Forecasting ERA5 reanalysis dataset (0.25 x 0.25 resolution). Values of these meteorological variables were determined at the village-level by identifying the four surrounding grid points with values available from the ERA5 reanalysis, and then applying inverse distance weighted interpolation from those four grid points to the village. Percent relative humidity was calculated from the 2-m dew point temperature using the "weathermetrics" package (version

Table 9: PMF error estimation diagnostics

	Potential Factor Solution			
Diagnostic	3	4	5	6
Qexp	27936	26052	24168	22284
Qtrue	187681	147796	123236	100316
Qrobust	174407	139910	117082	95932.5
Qr/Qexp	6.24	5.37	4.84	4.3
Q/Qexp > 6	wi-Ca, ns-S, ws-Na, ws-Ca, Al, Cl, Pb	ns-S, Na, Al, Cl, Pb, Nitrate	Nitrate, ws-Na, Al, Chloride	Nitrate, ws-Na, Al
DISP % dQ	<0.1%	<0.1%	<0.1%	<0.1%
DISP swaps	0	0	0	0
BS_mapping	Dust- 98.5%	Transported dust- 95%, Dust- 96.5%, Sulfur secondary- 97.5%, Mixed combustion- 96.5%	Transported dust- 86%, Mixed combustion- 87%, Dust- 86%, Lead- 55%	Transported dust- 84%, Mixed combustion- 87.5%, Dust- 81.5%, Lead- 72% Chloride- 61.5% Sulfur secondary- 98.5%

1.2.2) in R. Total hourly wind speed and wind direction were calculated from the horizontal wind speed components.

The model diagnostics for the three- to six-factor PMF solutions are given in Table 9. Model fit was assessed using Q/Qexp (how our model fit divided by the expected fit). As the change in Q/Qexp decreases as more factors are added, the model may be fitting additional sources that do not improve the overall fit. The largest change in Q/Qexp was from three to four sources (6.24 to 5.37) while the changes moving from four to five and five to six were similar, which suggests that four factors is sufficient and parsimonious to explain the variation in our data. We assessed the random error in our model by randomly sampling blocks of data, fitting new models with the blocks, and comparing how the source profiles compared to that of the original model (bootstrap (BS) mapping). The three- and four-factor solutions had high BS mapping (all factors found in $> 96.5\%$ of bootstrap runs). The additional sources identified in the five-factor (lead) and six-factor (chloride) solutions have low BS mapping ($> 72\%$), which means those solutions are not as consistent as the three- and four-factor solutions. The possibility that multiple, different, solutions could result in the same Q value was assessed using displacement. The displacement approach takes the original factor profiles and modifies the values for each species up or down to maintain a small change in Q, reruns the solution with the new species values, and then compares the profiles of the new model to the original. Any swaps indicate that small changes in the species values could result in factor profiles that look different from the original solution, and that the original solution is unstable. None of the factors in any of the solutions discussed were swapped during displacement, which indicates that all of the potential solutions are stable. Based on the Q/Qexp, BS mapping, and interpretability of the factors, the four-factor solution was selected as the most appropriate source solution for the data.

The source profiles for the four-factor solution are presented in Figure 6. The first source was identified as dust based on high percentages of crustal elements like wi-Ca, Si, and wi-Mg. The second source consisted of non-sulfate sulfur as well as secondary inorganic ions (ammonium, nitrate, and sulfate). Non-sulfate sulfur is a tracer for primary coal combustion, while secondary inorganic ions indicate a secondary source. Since industrial coal burning is a source of power generation in our study area, it is likely that the second source is a mixture of primary and secondary emissions that originate from coal and other sulfurous fuel combustion. Additionally, the mean source contribution of the second source is higher in outdoor than personal exposure measurements. Secondary formation occurs outdoors in the presence of sunlight, so higher outdoor concentrations compared to personal exposure further support our naming the second source 'sulfur secondary'. The third source had high percentages of ws-Ca nd Al, which in our study region, has been found to be indicative of transported dust from dust storms that can occur in the spring. While our samples were collected during winter months only, it is possible that transported dust from previous years still remained. The fourth source was characterized by high percentages of tracers for both coal (OC, wi-K, chloride, Pb) and biomass combustion (EC, ws-K). Coal and biomass combustion are anticipated sources of PM pollution in our study setting, particularly from domestic cooking and heating activities, so this source is likely a mixture of PM emitted from these two household combustion sources.

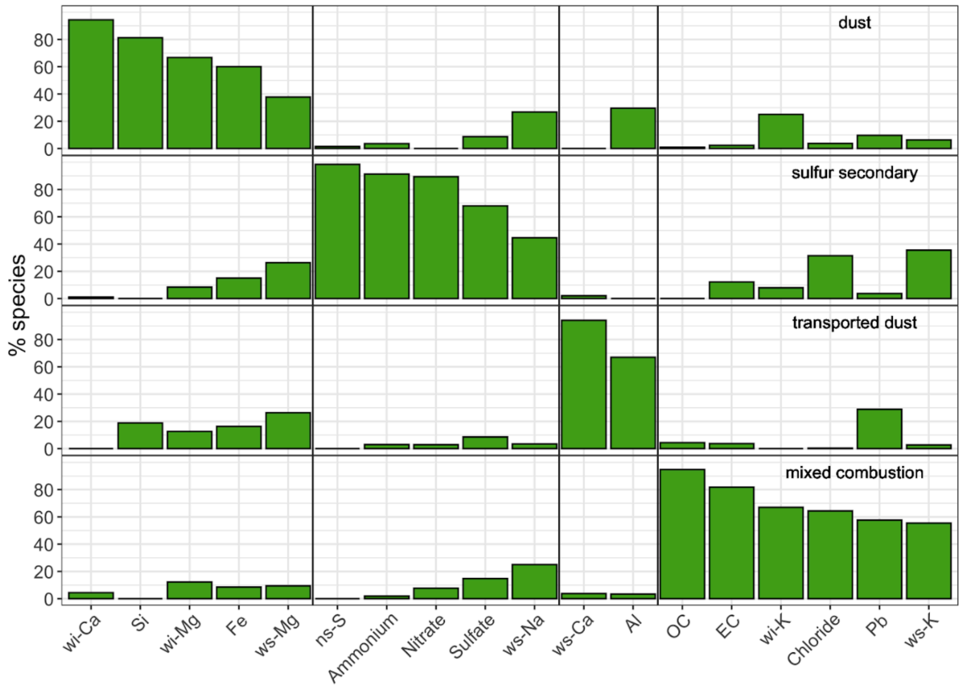


Figure 6: Source profiles for the 4-factor PMF solution to the sum of elements, ions, elemental carbon, and organic carbon for outdoor and personal $PM_{2.5}$ exposure measurements. The lines separate the major contributing species to each source

We extend the source profiles across the different treatment cohorts in Figure 7.

Table 10: Average treatment effect (μg/m^3) for outdoor and personal exposure to the mixed combustion source.

	DiD		Adjusted DiD[a]	
	ATT	(95% CI)	ATT	(95% CI)
Outdoor	1.07	(-4.90, 7.04)	1.53	(-4.19, 7.26)
Personal exposure	-5.60	(-13.70, 2.54)	-5.39	(-13.1, 2.35)

[a] Note: Models adjusted for ???

Figure 7: Arithmetic mean dispersion normalized source contributions found from the 4-factor PMF solution for A outdoor and B personal $PM_{2.5}$ exposure samples by year the group received treatment.

Overall, Table 10 shows the average treatment effect of the CBHP policy on outdoor (community-level) and personal exposure to the mixed combustion source was statistically indistinguishable from the null. Treatment was associated with lower, but statistically imprecise, personal exposures to the mixed combustion source. As with BC, this finding is consistent with the expectation that the policy contributed to reducing air pollutant emissions from solid fuel burning, as this 'mixed combustion' source most likely reflects solid fuel combustion, particularly in our study settings. The results were consistent in the unadjusted and adjusted models.

When the average treatment effect of the CBHP policy on outdoor (community-level) and personal exposure to the mixed combustion source was were allowed to vary by treatment year and time, the treatment effect for households most recently treated (i.e., treated in the final season, Season 4) was associated with lower personal exposures to the mixed combustion source (Appendix Figure A5). In each season, treatment by the CBHP policy was associated with a reduction in the source contribution to personal $PM_{2.5}$ mass from the mixed combustion source; however, for villages treated in Seasons 2 and 3, the effect was statistically imprecise. Treatment was not associated with a reduction or an increase in the source contribution to community outdoor $PM_{2.5}$ mass from the mixed combustion source. That personal exposure measures of this specific air pollution source were more indicative of treatment effect than community outdoor measures of the same air pollution source is consistent with the expectation that this source, which we determined to be indicative of a mixture of coal and biomass combustion, is most characteristic of household use of solid fuels, including coal and biomass, which would produce emissions that are likely to be nearer to the people using those fuels than near to the centrally located community outdoor air samplers.

6.6 Aim 3: Mediation by source contribution

Table 11 shows results from the mediation analysis by personal exposure to the mixed combustion source (coal and biomass), estimated for the subset of participants with personal exposure measurements. The CDE in this model estimates the impact of exposure to the CBHP policy on central and systolic blood pressure while holding constant values of mixed combustion source at the mean baseline values for untreated population. The marginal policy effects (ATTs) from the adjusted ETWFE models for this subset of participants were largely similar to those from the full sample for central SBP (around a 1.6 mmHg decrease), but slightly smaller for central DBP (-1.7 mmHg in the full sample vs. -1.3 in the subset with personal exposure measurements) and were estimated with greater imprecision. We found little evidence that these treatment effects were meaningfully mediated by exposure to the mixed combustion source, as the controlled direct effects were generally of similar magnitude as the adjusted total effects.

7 Discussion and Conclusions

Air pollution emitted from residential space heating with coal has historically been a major contributor to cardio-respiratory disease burden in northern China (Archer-Nicholls et al. 2016; Yun et al. 2020). Since the introduction of its 13th 5-Year-Plan (2016-2020), China has successfully implemented numerous large-scale measures to improve air quality including programs that incentivize rural household transition from solid fuels to clean energy sources (Young et al. 2015). The CBHP policy is among the largest and most ambitious household energy policies implemented anywhere in the world in recent decades, and its staggered roll-out provided us with a unique opportunity to prospectively evaluate this real-world experiment and its effects on air quality and health.

Table 11: Average treatment effects and controlled direct effect (mm/Hg) of the CBHP policy on central systolic and diastolic blood pressure with mixed combustion source as the potential mediator.

	DiD		Adjusted DiD[a]		Adjusted CDE[b]	
	ATT	(95%CI)	ATT	(95%CI)	ATT	(95%CI)
Central SBP	-0.43	(-3.48, 2.62)	-1.61	(-5.03, 1.82)	-1.5	(-4.94, 1.94)
Central DBP	-0.58	(-2.59, 1.43)	-1.26	(-3.38, 0.87)	-1.37	(-3.69, 0.96)

Note: ATT = Average Treatment Effect on the Treated, DiD = Difference-in-Differences, CDE = Controlled Direct Effect, DBP = Diastolic Blood Pressure, SBP = Systolic Blood Pressure.

[a] Adjusted for age, sex, waist circumference, smoking, alcohol consumption, and use of blood pressure medication.

[b] Further adjusted for mediation by mixed combustion source (coal and biomass)

7.1 Adoption of the heat pump technology and adherence to the policy

The CBHP policy was successful in driving a rapid household heating energy transition from coal to electric heating in our treated villages. There was high uptake and consistent use of the new heat pump technology and large reductions in coal use in treated villages starting in the first year post-treatment and continuing into the third year for villages treated in 2019. We enrolled rural and peri-urban villages across a wide geographic area and socioeconomic spectrum in Beijing and observed near universal adoption of the heat pump technologies and suspension of coal stove use across the different treatment groups and campaigns. This contrasts with many previous household energy intervention studies, including several randomized trials, where low fidelity and compliance with the intervention were considered a major limitation to achieving their intended air quality or health benefits (Ezzati and Baumgartner 2017; Harrison et al. Approved February 2024; Rosenthal et al. 2018).

A number of factors contribute to the successful uptake of the new technology and adherence to the policy. The initial uptake of the heat pump was influenced by broad support and perceived benefits of village and household participation in the policy. At baseline assessment, 49 of 50 village committee interviewees indicated a desire to participate in the policy by the committee members and their constituents, for reasons including ease of use of the heat pump and the convenience of no longer having to add coal throughout the day and night, the desire for a cleaner local environment, and perceived lower risk of carbon monoxide poisoning compared with coal stoves. While the availability and cost of clean fuels were barriers to clean fuel adoption in previous studies (Rehfuess et al. 2014), in our study, both the upfront costs of the heat pump and electricity were heavily subsidized, which limited the financial burden of transition for households. Post-policy implementation, treated villages no longer had access to government-subsidized coal, and coal burning was further discouraged with possible punitive measures (e.g., potential loss of electricity subsidies).

7.2 Impacts of the policy on health

One of the key findings from our comprehensive evaluation of the CBHP policy was that exposure to the policy reduced systolic and diastolic blood pressure by ~1.5 mmHg, and that most of the observed BP effects were mediated by improvements in the indoor environment, specifically reductions in indoor $PM_{2.5}$ and increases in indoor temperature. The total effects of the policy are supported by a small number of randomized trials of cooking gas or more efficient biomass cookstoves showing similar or larger reductions in blood pressure (SBP: -2.1 to -1.3 mmHg; DBP: -0.1 to -3.0 mmHg) (Kumar et al. 2021). In contrast, a recent multi-country randomized trial of an LPG stove observed a small (~0.6 mmHg) increase in gestational blood pressure (Ye et al. 2022) despite decreases in exposure to $PM_{2.5}$ that were much larger than in our study. Though, the study participants were much younger than our study (mean age of 25 versus 61y) and gas stoves can still emit health-damaging air pollutants like benzine and volatile organic compounds (Kashtan et al. 2023), especially in contrast with the zero-emission electric heaters introduced in our study villages. Further, our findings of temperature- and air quality-mediated impacts of the policy on BP are supported by observational studies conducted by us and others showing that increased exposure to household air pollution (Baumgartner et al. 2018, 2011; Dong et al. 2013; Kanagasabai et al. 2022) and to colder indoor temperatures (Lv et al. 2022; Sternbach et al. 2022) are associated with higher blood pressure in rural and peri-urban areas of China, with exposure-response estimates that align with our mediator estimates.

We did not observe effects of the policy on blood pressure measures of PP or cPP/SBP amplification. Pulse pressure is measured as the difference between SBP and DBP, and represents the pulsatile component of blood flow (Dart and Kingwell 2001). Thus, increases in PP can result from increases in SBP, decreases in DBP, or both. The lack of effect on PP in our study is likely attributed to the near identical reductions in SBP and DBP from the policy. Similarly, PP/SBP amplification is measured as a ratio of peripheral to central pressures, and the decreases in central and brachial pressures with the policy were also nearly identical in our study. Although the duration of our study was nearly twice as long as most previous household stove intervention studies, it is still possible that longer-term reductions in BP are required to observe any structural changes in the caliber or elasticity of arterial walls that would be reflected in differences in measures PP or SBP/PP amplification (Dart and Kingwell 2001).

Exposure to the CBHP policy also reduced self-report of any poor respiratory symptoms (~7 percentage points) with most of these effects driven by reductions in self-reported chest trouble or difficulty breathing on several or most days of the week. In Guatemala exposure to carbon monoxide (used as a surrogate for exposure to $PM_{2.5}$) and prevalence of chronic respiratory symptoms, especially wheeze, was reduced among women who received a biomass chimney stove (i.e., plancha) (Smith-Sivertsen et al. 2009). [[[Note to Jill: still need to review and discuss results from trials with respiratory symptoms: Fandino-Del-Rio et al., 2022, Romieu et al., 2009; Schilmann et al., 2015, Add in null findings from Burwen et al., 2012, Beltramo et al.,2013]]]

We found some evidence of heterogeneity in the health benefits of the policy by treatment cohort, specifically a small increase in BP (2.4 mmHg; 95%CI: -0.5, 5.3) and self-reported wheeze events (+9 percentage points; 95%CI: 0, 18) in the three villages treated in 2021. Notably this is also the treatment cohort with the smallest decline in point temperature at the time of BP measurement and both an increase in self-reported biomass use and several households that continued using coal. Rather paradoxically, we observed a larger decrease in $PM_{2.5}$ and mixed solid fuel use in this group. It is possible that the composition of PM and mixed solid fuel was different in this cohort, with a greater contribution of biomass smoke, however we are unable to differentiate between biomass and coal in our 'mixed solid fuel' category. This group was also treated during the pandemic, which could have impacted how the policy was introduced or resulted in changes to other BP risk factors that we did not evaluate in our study, e.g., changes in diet or level of social capital.

We also did not observe impacts of the policy on blood biomarkers of inflammation and oxidative stress in the sub-sample of participants with blood collection in S1 and S2. Our results contrast with a natural experiment in urban Beijing that showed large regional and local air quality reductions during the 2008 Beijing Olympics and also observed benefits to airway inflammation (Huang et al. 2012) and blood markers of inflammation and oxidative stress in healthy urban Beijing residents during the Olympics compared with before and after (Rich et al. 2012). Our mediation analysis indicated that the blood pressure effects of the policy were mediated more through indoor temperature than air pollution. Although observational studies from rural northern China do show impacts of exposure to temperature on inflammation and oxidative stress (Wang et al. 2020; Xu et al. 2019), it's possible that the relatively small increases in mean indoor temperature in treated households were not sufficiently large to capture measurable changes in these biomarkers.

7.3 Impacts of the policy on air pollution and is sources

China has a long history of launching ambitious, large-scale policies and programs to promote clean household energy transition and support rural energy infrastructure development (Zhang and Smith 2007). The country was a relatively early initiator of rural electrification projects in the 1950s and achieved complete (100%) electrification of households by 2016 (Yang 2021), which undoubtedly facilitated the policy choice to replace coal stoves with electric-powered heat pump heaters. Several decades earlier, China achieved what is likely the largest improvement in energy efficiency in history in terms of the population affected by just one program. The National Improved Stove Program (NISP) and its provincial counterparts were initiated in the early 1980s and are credited with introducing nearly 200 million improved cooking and heating stoves by the late 1990s. The NISP implemented mostly chimney stoves with the primary goal of increased fuel efficiency to reduce pressure on local forests and a secondary goal of reducing indoor pollution. NISP biomass stoves showed some success in reducing indoor PM, though levels were still much higher than health-motivated guidelines, however the program's so called 'improved' coal stoves provided no measurable air quality benefit (Sinton et al. 2004).

In contrast, our evaluation of the CBHP policy revealed a quantifiable improvement in indoor air quality, evidenced by a reduction of 36 µg/m^3 in seasonal measures of indoor $PM_{2.5}$. Compared to the current World Health Organization (WHO) guidelines, which state that annual (and 24-h) average concentrations of $PM_{2.5}$ should not exceed 5 (and 15) µg/m^3 (Organization 2021), homes at baseline in this study had indoor $PM_{2.5}$ in the range of interim targets (ITs) 1 and 2. Interim targets have been set by WHO to support the planning of incremental milestones toward cleaner air, particularly for cities, regions, and countries with higher air pollution levels. For $PM_{2.5}$, the four interim (IT) targets for annual and 24-h means are: IT-1: 35 and 75 µg/m^3; IT-2: 25 and 50 µg/m^3; IT-3: 15 and 37.5 µg/m^3; and IT-4: 10 and 25 µg/m^3 (Organization 2021). Indoor $PM_{2.5}$ concentrations were reduced in treated homes, bringing these homes into the range of IT-4[b] and realizing some of the considerable potential for air quality improvement from the intervention.

The still elevated levels of indoor and personal exposures in treated homes, despite high compliance with the policy, is most likely attributable to two main factors: the elevated levels of ambient air pollution in our study setting (range: 26-38 g/m^3 in treated villages) and the continued use of inefficient and highly-polluting biomass-burning kangs. Kangs are a relatively simple and culturally entrenched space heating technique that has been used for over two thousand years in China (Zhuang et al. 2009). Kangs are usually fuelled by wood or other biomass that is freely and readily available in our rural and peri-urban study villages. The CBHP policy did not ban biomass burning, and we did observe persistent use of kangs based on both household surveys and the $PM_{2.5}$ source analysis and apportionment results. The continued use of solid fuel stoves (i.e., stove stacking) alongside cleaner stoves and fuels has long been a barrier to achieving the intended air quality benefits of household energy interventions (Shankar et al. 2020). A notable exception is a recently completed multi-country randomized trial of LPG cookstoves which attained near exclusive use of LPG stoves and dramatic reductions in personal exposures to $PM_{2.5}$ (lowered by 66% compared with controls) (Johnson et al. 2022), but rather unexpectedly did not observe health benefits across a range of neonatal, child, and maternal outcomes (Harrison et al. Approved February 2024).

In addition to the Household Air Pollution Intervention Network (HAPIN) trial mentioned above, several other recent trials (Alexander et al. 2017; Checkley et al. 2021; Chillrud et al. 2021; Katz et al. 2020) of interventions have evaluated the impacts of replacing solid fuels for cooking with cleaner fuels such as liquefied petroleum gas (LPG) or ethanol. These trials have reported reductions in in-home measures of air pollutants indicative of combustion (i.e., PM, carbon monoxide). Yet, baseline pollutant concentrations tended to be higher than what we observed in our study, in some cases exceeding several hundreds of micrograms of $PM_{2.5}$ per cubic meter of air (Rosa et al. 2014). Thus, even higher percentage reductions in air pollution levels than what we observed still left WHO interim target levels out of reach. It may also be worth noting that differences in summary metrics makes comparisons between our study and others challenging; compared to other household air pollution intervention studies, our study yielded a larger sample size of longer-duration, repeated measures of indoor air pollution (Thomas et al. 2015).

Despite the significant improvements in indoor air quality, the intervention's effects on reducing personal exposure to $PM_{2.5}$, BC, and source contributions from the 'mixed combustion' source identified through our study's source apportionment, as well as ambient outdoor levels of the same air pollution measures, were more limited. The limited effect of the intervention on personal and outdoor $PM_{2.5}$ and BC levels suggests the persistent influence of pollution sources that are likely external to the home environment (e.g., vehicular emissions, industrial activities, power generation). The source analysis and apportionment identified a source of mixed combustion as a significant contributor to $PM_{2.5}$, implying that, in addition to space heating solid fuel combustion, non-heating related combustion activities may have also persisted in their contribution to personal exposure and outdoor pollution.

It may be valuable to compare our results to several recent studies that were not focused on household air pollution interventions, but still measured indoor, personal, and / or outdoor air pollution in settings that might be influenced by similar sources to ours…

In this study we were best able to quantify air pollution impacts from the CBHP policy using long-term measurements at locations nearest to indoor stoves, which required hundreds of air monitors and thousands of hours of measurements. The scale and duration of air pollution measurement achieved in this study would not have been possible without low-cost air pollution sensors that have proliferated in the past decade. [[[Ellison to add one or two relevant references from China-based sensor studies, and maybe one or two reviews of the low-cost sensor transformation of air pollution measurement]]]. Evaluation methods for future intervention studies might also include longitudinal measures of air pollution that track changes over longer periods to capture delayed effects. An examination of how we evaluate the effectiveness of interventions like the CBHP policy raises fundamental questions about the adequacy of our current approaches. Traditional air pollution metrics may not fully capture the broader, systemic changes that such policies aim to achieve.

Determining whether the CBHP policy worked required a multifaceted approach to evaluation that went beyond measures of air quality alone. By incorporating a broader array of metrics and considering the systemic nature of air pollution and its health impacts, through this study, we sought to provide a more nuanced understanding of an intervention's effectiveness and the ways in which it may need to be augmented or restructured to achieve desired health outcomes.

7.4 Assumptions, strengths and limitations

The validity of our DiD approach is subject to several key assumptions. First, we assume that anticipation of the CBHP policy did not differ between treated and untreated villages. We selected villages that were eligible for the policy but not currently treated. It was generally understood that the policy would first be implemented in the plains areas with more updated electric grids and then gradually expand into more remote and mountainous areas of Beijing, though most of our study villages were far from Beijing's urban core. In addition to these geographical parameters, some of our study villages were assigned to the policy whereas others applied to the local government, but they were generally unaware if or when they would be treated at the time of enrollment. Second, our analysis assumes that, in the absence of the policy, the trends in air quality and health in treated and untreated villages would have remained the same over time. While we cannot fully verify this assumption, we observed similar trends in health outcomes between S1 and S2 in never treated villages and those treated later in S3 or S4 (ref[c] SI figures that Talia created a while back for BP and resp outcomes) and we adjusted for relevant time-varying confounders, all of which improve the credibility of this assumption. Fourth, we cannot entirely rule out the possibility that other programs or policies differentially affected air quality or health in treated and untreated villages, which could lead to over- or under-estimation of its effects. Though, we surveyed village leaders about other rural development or health policies and programs in their villages throughout our four-year study period and did not identify any co-implemented programs that would differentially impact treatment groups. Finally, our mediation analysis assumes no residual confounding between our mediators (air pollution and temperature) and our health outcomes. [[[need to add text]]]

Strengths[d] of this comprehensive, field-based assessment of the CBHP policy include our quasi-experimental design to evaluate a real-world clean energy intervention that would be near impossible to experimentally manipulate at the scale of our study. Our study design controlled for secular changes in health and we additionally collected data on and adjusted for important time-varying covariates. Our numerous sensitivity analyses showed the robustness of our findings to various analytic decisions. Most previous field-based household energy intervention studies were less than two-years in duration (Harrison et al. Approved February 2024; Quansah et al. 2017), and our four-year study enabled longer-term evaluation of compliance with the coal ban and heat pump adoption/use and their impacts on air pollution and health. We retained all 50 villages in this assessment of a village-level intervention, and were able to successfully obtain measurements from over 1000 participants into each campaign with individual measurements despite half of our study occurring during the covid-19 pandemic. By comparison, previous field-based assessments of household energy interventions (trials and pre-post designs with controls) and blood pressure ranged in size from 44 to 324 participants (Harrison et al. Approved February 2024; Kumar et al. 2021; Onakomaiya et al. 2019), with exception of the recent multi-country trial that enrolled ~3000 pregnant women (Ye et al. 2022).

This study also has several limitations to consider when interpreting our results.

First, the covid pandemic started in the middle of our study and roughly half of our treated villages went into the policy during the pandemic, which likely had some influence on the roll-out of the policy in those villages. We observed the largest benefits in BP and several respiratory outcomes in villages treated before the pandemic. However, we cannot differentiate between treatment cohort effects attributable the covid pandemic versus other factors that different between treatment cohorts (i.e., geographic location, access to biomass, fuel prices, etc)

Second, the policy roll-out began in 2016 though we did not begin enrolling villages into our study until 2018. Thus, our study villages are farther from the urban core and generally of lower SES than many villages treated in the first three years of the policy. Previous studies of the CBHP policy suggest that treated villages of all SES levels benefited from less-polluted and warmer indoor environments, but that the benefits were smaller in lower SES villages compared with higher SES villages (Barrington-Leigh et al. 2019; Meng et al. 2023). Further, our more remote and rural study villages also have easier access to biomass fuel. Thus our results may not be generalizable to all of Beijing, especially the more urbanized, higher SES villages treated between 2016 and 2018, and may underestimate the impacts of the policy on indoor environmental factors that were important cardio-respiratory health mediators in our study.

Third, like any field-based study, we had a number of constraints with data collection. We were unable to measure indoor air quality or stove use in S1 due to logistical and budget constraints, and thus cannot directly estimate the effects of the policy on indoor PM2.5 for the 10 villages treated in 2019. Similarly, we were unable to take blood in the last campaign because measurements were in homes rather than clinics to avoid group contact during the pandemic. In addition, our study logistics required visiting 50 villages over a period of just several months. Thus, we were unable to return to villages if a previously enrolled participant was not at home at the time that staff visited the village. In such instances, we either randomly selected either another eligible participant in the same home or we randomly selected another household with eligible participants from the village roster and our study participants different slightly across campaigns. Our village-level study and analysis is robust to participation of a random sample of participants in each campaign, and we also found no notable differences in key demographic characteristics or health behaviors between participants who contributed to a different number of campaigns or between participants across each of the three campaigns.

Fourth, our respiratory symptoms are self reported and thus our estimated effects of the policy must be interpreted with caution given that participants are not blinded to the household energy intervention. In previous studies of household water filters, for example, (Peel et al. 2015)

8 Implications of Findings

In this comprehensive field-based assessment of the CBHP policy in Beijing, we observed high fidelity and compliance with the policy in our study villages and households where nearly all households in treated villages stopped using coal and shifted to electric-powered heaters. Exposure to the policy reduced blood pressure and self-reported chronic respiratory symptoms, and the effects were mediated by reductions in indoor $PM_{2.5}$ and improvements in home temperature. We did not observe the same benefits of the policy on outdoor air quality or personal exposures, likely because the relatively high contribution of other regional and local air pollution sources to outdoor and personal exposures may have masked the benefits from a single source reduction. We also did not observe benefits of the policy on different measures of inflammation and oxidative stress in the subsample of participants with biomarker assessment, even though we observed respiratory symptoms and BP benefits of the policy in a sensitivity analysis limited to the same participants. [e]

Our results showing a home environment and health benefit of a large-scale and successfully implemented clean energy policy are timely, as they are synchronous with ongoing and planned clean energy policies in China and other countries in a global effort to "ensure access to affordable, reliable, sustainable, and modern energy for all" (Sustainable Development Goal-7) and also directly respond to a recent call-to-action from global cardiovascular societies that emphasized the urgent need for interventional studies that inform targeted pollution-reducing strategies to reduce cardiovascular disease (Brauer et al. 2021).

9 Data Availability Statement

- Description of datasets and code available on our project page at the Open Science Foundation

10 Acknowledgements

To come…

11 References

Ahmed T, Dutkiewicz VA, Shareef A, Tuncel G, Tuncel S, Husain L. 2009. Measurement of black carbon (BC) by an optical method and a thermal-optical method: Intercomparison for four sites. Atmospheric Environment 43:6305–6311; doi:10.1016/j.atmosenv.2009.09.031.

Alexander DA, Northcross A, Karrison T, Morhasson-Bello O, Wilson N, Atalabi OM, et al. 2018. Pregnancy outcomes and ethanol cook stove intervention: A randomized-controlled trial in Ibadan, Nigeria. Environment International 111:152–163; doi:10.1016/j.envint.2017.11.021.

Alexander D, Northcross A, Wilson N, Dutta A, Pandya R, Ibigbami T, et al. 2017. Randomized Controlled Ethanol Cookstove Intervention and Blood Pressure in Pregnant Nigerian Women. American Journal of Respiratory and Critical Care Medicine 195:1629–1639; doi:10.1164/rccm.201606-1177OC.

An L, Hong B, Cui X, Geng Y, Ma X. 2021. Outdoor thermal comfort during winter in China's cold regions: A comparative study. Science of The Total Environment 768:144464; doi:10.1016/j.scitotenv.2020.144464.

Archer-Nicholls S, Carter E, Kumar R, Xiao Q, Liu Y, Frostad J, et al. 2016. The Regional Impacts of Cooking and Heating Emissions on Ambient Air Quality and Disease Burden in China. Environmental Science & Technology 50:9416–9423; doi:10.1021/acs.est.6b02533.

Barrington-Leigh C, Baumgartner J, Carter E, Robinson BE, Tao S, Zhang Y. 2019. An evaluation of air quality, home heating and well-being under Beijing's programme to eliminate household coal use. Nature Energy 4:416–423; doi:10.1038/s41560-019-0386-2.

Baumgartner J, Carter E, Schauer JJ, Ezzati M, Daskalopoulou SS, Valois M-F, et al. 2018. Household air pollution and measures of blood pressure, arterial stiffness and central haemodynamics. Heart 104:1515–1521; doi:10.1136/heartjnl-2017-312595.

Baumgartner J, Schauer JJ, Ezzati M, Lu L, Cheng C, Patz JA, et al. 2011. Indoor Air Pollution and Blood Pressure in Adult Women Living in Rural China. Environmental Health Perspectives 119:1390–1395; doi:10.1289/ehp.1003371.

Brauer M, Casadei B, Harrington RA, Kovacs R, Sliwa K, the WHF Air Pollution Expert Group. 2021. Taking a Stand Against Air Pollution—The Impact on Cardiovascular Disease: A Joint Opinion From the World Heart Federation, American College of Cardiology, American Heart Association, and the European Society of Cardiology. Circulation 143; doi:10.1161/CIRCULATIONAHA.120.052666.

Callaway B. 2020. Difference-in-Differences for Policy Evaluation. In: *Handbook of Labor, Human Resources and Population Economics* (K.F. Zimmermann, ed). Springer International Publishing:Cham. 1–61.

Callaway B, Sant'Anna PHC. 2021. Difference-in-Differences with multiple time periods. Journal of Econometrics 225:200–230; doi:10.1016/j.jeconom.2020.12.001.

Card D, Krueger AB. 1994. Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. American Economic Review 84: 772–93.

Checkley W, Williams KN, Kephart JL, Fandiño-Del-Rio M, Steenland NK, Gonzales GF, et al. 2021. Effects of a Household Air Pollution Intervention with Liquefied Petroleum Gas on Cardiopulmonary Outcomes in Peru. A Randomized Controlled Trial. American Journal of Respiratory and Critical Care Medicine 203:1386–1397; doi:10.1164/rccm.202006-2319OC.

Chillrud SN, Ae-Ngibise KA, Gould CF, Owusu-Agyei S, Mujtaba M, Manu G, et al. 2021. The effect of clean cooking interventions on mother and child personal exposure to air pollution: Results from the Ghana Randomized Air Pollution and Health Study (GRAPHS). Journal of Exposure Science & Environmental Epidemiology 31:683–698; doi:10.1038/s41370-021-00309-5.

Clark S, Carter E, Shan M, Ni K, Niu H, Tseng JTW, et al. 2017. Adoption and use of a semi-gasifier cooking and water heating stove and fuel intervention in the Tibetan Plateau, China. Environmental Research Letters 12:075004; doi:10.1088/1748-9326/aa751e.

Costello BT, Schultz MG, Black JA, Sharman JE. 2015. Evaluation of a Brachial Cuff and Suprasystolic Waveform Algorithm Method to Noninvasively Derive Central Blood Pressure. American Journal of Hypertension 28:480–486; doi:10.1093/ajh/hpu163.

Dai Q, Liu B, Bi X, Wu J, Liang D, Zhang Y, et al. 2020. Dispersion Normalized PMF Provides Insights into the Significant Changes in Source Contributions to $PM_{2.5}$ after the COVID-19 Outbreak. Environmental Science & Technology 54:9917–9927; doi:10.1021/acs.est.0c02776.

Dart AM, Kingwell BA. 2001. Pulse pressure—a review of mechanisms and clinical relevance. Journal of the American College of Cardiology 37:975–984; doi:10.1016/S0735-1097(01)01108-

1.

Dispersed Coal Management Research Group . 2023. China Dispersed Coal Governance Report.

Dockery DW, Rich DQ, Goodman PG, Clancy L, Ohman-Strickland P, George P, et al. 2013. Effect of air pollution control on mortality and hospital admissions in Ireland. Research Report (Health Effects Institute) 3–109.

Dominici F, Greenstone M, Sunstein CR. 2014. Science and regulation. Particulate matter matters. Science (New York, NY) 344:257–9; doi:10.1126/science.1247348.

Dong G-H, Qian Z(Min), Xaverius PK, Trevathan E, Maalouf S, Parker J, et al. 2013. Association Between Long-Term Air Pollution and Increased Blood Pressure and Hypertension in China. Hypertension 61:578–584; doi:10.1161/HYPERTENSIONAHA.111.00003.

Duan X, Jiang Y, Wang B, Zhao X, Shen G, Cao S, et al. 2014. Household fuel use for cooking and heating in China: Results from the first Chinese Environmental Exposure-Related Human Activity Patterns Survey (CEERHAPS). Applied Energy 136:692–703; doi:10.1016/j.apenergy.2014.09.066.

Ezzati M, Baumgartner JC. 2017. Household energy and health: Where next for research and practice? Lancet (London, England) 389:130–132; doi:10.1016/S0140-6736(16)32506-5.

Food and Drug Administration. 2018. Bioanalytical Method Validation Guidance for Industry.

Goin DE, Riddell CA. 2023. Comparing Two-way Fixed Effects and New Estimators for Difference-in-Differences: A Simulation Study and Empirical Example. Epidemiology 34:535; doi:10.1097/EDE.0000000000001611.

Goodman-Bacon A. 2021. Difference-in-differences with variation in treatment timing. Journal of Econometrics 225:254–277; doi:10.1016/j.jeconom.2021.03.014.

Gould CF, Bejarano ML, Kioumourtzoglou M-A, Lee AG, Pillarisetti A, Schlesinger SB, et al. 2023. Widespread Clean Cooking Fuel Scale-Up and under-5 Lower Respiratory Infection Mortality: An Ecological Analysis in Ecuador, 1990–2019. Environmental Health Perspectives 131:037017; doi:10.1289/EHP11016.

Group GMW. 2016. Burden of disease attributable to coal-burning and other air pollution sources in China.

Harrison K, Hartinger S, Jack D, Kaali S, Lydston M, Mortimer K, et al. Approved February 2024. Household Air Pollution Interventions to Improve Health in Low- and Middle-Income Countries: An Official American Thoracic Society Research Statement.

Huang W, Wang G, Lu S-E, Kipen H, Wang Y, Hu M, et al. 2012. Inflammatory and Oxidative Stress Responses of Healthy Young Adults to Changes in Air Quality during the Beijing Olympics. American Journal of Respiratory and Critical Care Medicine 186:1150–1159; doi:10.1164/rccm.201205-0850OC.

International R. 2009. Standard Operating Procedure for the X-Ray Fluorescence Analysis of Particulate Matter Deposits on Teflon Filters: PM Xrf Analysis.

Johnson M, Pillarisetti A, Piedrahita R, Balakrishnan K, Peel JL, Steenland K, et al. 2022. Exposure Contrasts of Pregnant Women during the Household Air Pollution Intervention Network Randomized Controlled Trial. Environmental Health Perspectives 130:097005;

doi:10.1289/EHP10295.
Johnston FH, Hanigan IC, Henderson SB, Morgan GG. 2013. Evaluation of interventions to reduce air pollution from biomass smoke on mortality in Launceston, Australia: Retrospective analysis of daily mortality, 1994-2007. BMJ 346:e8446–e8446; doi:10.1136/bmj.e8446.
Kanagasabai T, Xie W, Yan L, Zhao L, Carter E, Guo D, et al. 2022. Household Air Pollution and Blood Pressure, Vascular Damage, and Subclinical Indicators of Cardiovascular Disease in Older Chinese Adults. American Journal of Hypertension 35:121–131; doi:10.1093/ajh/hpab141.
Kashtan YS, Nicholson M, Finnegan C, Ouyang Z, Lebel ED, Michanowicz DR, et al. 2023. Gas and Propane Combustion from Stoves Emits Benzene and Increases Indoor Air Pollution. Environmental Science & Technology 57:9653–9663; doi:10.1021/acs.est.2c09289.
Katz J, Tielsch JM, Khatry SK, Shrestha L, Breysse P, Zeger SL, et al. 2020. Impact of Improved Biomass and Liquid Petroleum Gas Stoves on Birth Outcomes in Rural Nepal: Results of 2 Randomized Trials. Global Health: Science and Practice 8:372–382; doi:10.9745/GHSP-D-20-00011.
Keele L, Tingley D, Yamamoto T. 2015. Identifying mechanisms behind policy interventions via causal mediation analysis. Journal of Policy Analysis and Management 34: 937–963.
Khuzestani RB, Schauer JJ, Wei Y, Zhang Y, Zhang Y. 2017. A non-destructive optical color space sensing system to quantify elemental and organic carbon in atmospheric particulate matter on Teflon and quartz filters. Atmospheric Environment 149:84–94; doi:10.1016/j.atmosenv.2016.11.002.
Kumar N, Phillip E, Cooper H, Davis M, Langevin J, Clifford M, et al. 2021. Do improved biomass cookstove interventions improve indoor air quality and blood pressure? A systematic review and meta-analysis. Environmental Pollution 290:117997; doi:10.1016/j.envpol.2021.117997.
Lai. 2019. Relative contributions of household solid fuel use and outdoor air pollution to chemical components of personal PM2.5 exposures. Indoor Air-international Journal of Indoor Air Quality and Climate.
Lewington S, LiMing L, Sherliker P, Yu G, Millwood I, Zheng B, et al. 2012. Seasonal variation in blood pressure and its relationship with outdoor temperature in 10 diverse regions of China: The China Kadoorie Biobank. Journal of hypertension 30: 1383.
Lindemann U, Stotz A, Beyer N, Oksa J, Skelton DA, Becker C, et al. 2017. Effect of indoor temperature on physical performance in older adults during days with normal temperature and heat waves. International journal of environmental research and public health 14; doi:10.3390/ijerph14020186.
Lowe A, Harrison W, El-Aklouk E, Ruygrok P, Al-Jumaily AM. 2009. Non-invasive model-based estimation of aortic pulse pressure using suprasystolic brachial pressure waveforms. Journal of Biomechanics 42:2111–2115; doi:10.1016/j.jbiomech.2009.05.029.
Lv Y, Zhu R, Xie J, Yoshino H. 2022. Indoor environment and the blood pressure of elderly in the cold region of China. Indoor and Built Environment 31:2482–2498; doi:10.1177/1420326X221109510.
McCracken JP, Smith KR, Díaz A, Mittleman MA, Schwartz J. 2007. Chimney Stove Intervention to Reduce Long-term Wood Smoke Exposure Lowers Blood Pressure among Guatemalan

Women. Environmental Health Perspectives 115:996–1001; doi:10.1289/ehp.9888.
McCracken J, Smith KR, Stone P, Díaz A, Arana B, Schwartz J. 2011. Intervention to Lower Household Wood Smoke Exposure in Guatemala Reduces ST-Segment Depression on Electrocardiograms. Environmental Health Perspectives 119:1562–1568; doi:10.1289/ehp.1002834.
Meng W, Zhu L, Liang Z, Xu H, Zhang W, Li J, et al. 2023. Significant but Inequitable Cost-Effective Benefits of a Clean Heating Campaign in Northern China. Environmental Science & Technology 57:8467–8475; doi:10.1021/acs.est.2c07492.
Naimi AI, Kaufman JS, MacLehose RF. 2014. Mediation misgivings: Ambiguous clinical and public health interpretations of natural direct and indirect effects. International journal of epidemiology 43:1656–61; doi:10.1093/ije/dyu107.
Niu J, Chen X, Sun S. 2024. China's Coal Ban policy: Clearing skies, challenging growth. Journal of Environmental Management 349:119420; doi:10.1016/j.jenvman.2023.119420.
Olson MR, Graham E, Hamad S, Uchupalanun P, Ramanathan N, Schauer JJ. 2016. Quantification of elemental and organic carbon in atmospheric particulate matter using color space sensing—hue, saturation, and value (HSV) coordinates. Science of The Total Environment 548–549:252–259; doi:10.1016/j.scitotenv.2016.01.032.
Onakomaiya D, Gyamfi J, Iwelunmor J, Opeyemi J, Oluwasanmi M, Obiezu-Umeh C, et al. 2019. Implementation of clean cookstove interventions and its effects on blood pressure in low-income and middle-income countries: Systematic review. BMJ Open 9:e026517; doi:10.1136/bmjopen-2018-026517.
Organization WH. 2021. WHO Global Air Quality Guidelines: Particulate Matter PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide.
Peel JL, Baumgartner J, Wellenius GA, Clark ML, Smith KR. 2015. Are Randomized Trials Necessary to Advance Epidemiologic Research on Household Air Pollution? Current Epidemiology Reports 2:263–270; doi:10.1007/s40471-015-0054-4.
Quansah R, Semple S, Ochieng CA, Juvekar S, Armah FA, Luginaah I, et al. 2017. Effectiveness of interventions to reduce household air pollution and/or improve health in homes using solid fuel in low-and-middle income countries: A systematic review and meta-analysis. Environment International 103:73–90; doi:10.1016/j.envint.2017.03.010.
Rehfuess EA, Puzzolo E, Stanistreet D, Pope D, Bruce NG. 2014. Enablers and Barriers to Large-Scale Uptake of Improved Solid Fuel Stoves: A Systematic Review. Environmental Health Perspectives 122:120–130; doi:10.1289/ehp.1306639.
Rich DQ, Kipen HM, Huang W, Wang G, Wang Y, Zhu P, et al. 2012. Association Between Changes in Air Pollution Levels During the Beijing Olympics and Biomarkers of Inflammation and Thrombosis in Healthy Young Adults. JAMA 307; doi:10.1001/jama.2012.3488.
Rosa G, Majorin F, Boisson S, Barstow C, Johnson M, Kirby M, et al. 2014. Assessing the Impact of Water Filters and Improved Cook Stoves on Drinking Water Quality and Household Air Pollution: A Randomised Controlled Trial in Rwanda. R.K. Hills, ed PLoS ONE 9:e91011; doi:10.1371/journal.pone.0091011.
Rosenthal J, Quinn A, Grieshop AP, Pillarisetti A, Glass RI. 2018. Clean cooking and the SDGs: Integrated analytical approaches to guide energy interventions for health and environment goals.

Energy for sustainable development : the journal of the International Energy Initiative 42:152–159; doi:10.1016/j.esd.2017.11.003.

Rubin DB. 1987. *Multiple Imputation for Nonresponse in Surveys*. 1st ed. Wiley.

Ruiz-Mercado I, Canuz E, Walker JL, Smith KR. 2013. Quantitative metrics of stove adoption using Stove Use Monitors (SUMs). Biomass and Bioenergy 57:136–148; doi:10.1016/j.biombioe.2013.07.002.

Scott AJ, Scarrott C. 2011. Impacts of residential heating intervention measures on air quality and progress towards targets in Christchurch and Timaru, New Zealand. Atmospheric Environment 45:2972–2980; doi:10.1016/j.atmosenv.2010.09.008.

Shang J, Zhang Y, Schauer JJ, Tian J, Hua J, Han T, et al. 2020. Associations between source-resolved PM2.5 and airway inflammation at urban and rural locations in Beijing. Environment International 139:105635; doi:10.1016/j.envint.2020.105635.

Shankar AV, Quinn AK, Dickinson KL, Williams KN, Masera O, Charron D, et al. 2020. Everybody stacks: Lessons from household energy case studies to inform design principles for clean energy transitions. Energy Policy 141:111468; doi:10.1016/j.enpol.2020.111468.

Sinton JE, Smith KR, Peabody JW, Yaping L, Xiliang Z, Edwards R, et al. 2004. An assessment of programs to promote improved household stoves in China. Energy for Sustainable Development 8:33–52; doi:10.1016/S0973-0826(08)60465-2.

Smith-Sivertsen T, Díaz E, Pope D, Lie RT, Díaz A, McCracken J, et al. 2009. Effect of Reducing Indoor Air Pollution on Women's Respiratory Symptoms and Lung Function: The RESPIRE Randomized Trial, Guatemala. American Journal of Epidemiology 170:211–220; doi:10.1093/aje/kwp100.

Snider G, Carter E, Clark S, Tseng J(TzuW, Yang X, Ezzati M, et al. 2018. Impacts of stove use patterns and outdoor air quality on household air pollution and cardiovascular mortality in southwestern China. Environment International 117:116–124; doi:10.1016/j.envint.2018.04.048.

Song C, Liu B, Cheng K, Cole MA, Dai Q, Elliott RJR, et al. 2023. Attribution of Air Quality Benefits to Clean Winter Heating Policies in China: Combining Machine Learning with Causal Inference. Environmental Science & Technology 57:17707–17717; doi:10.1021/acs.est.2c06800.

Sternbach TJ, Harper S, Li X, Zhang X, Carter E, Zhang Y, et al. 2022. Effects of indoor and outdoor temperatures on blood pressure and central hemodynamics in a wintertime longitudinal study of Chinese adults. Journal of Hypertension 40:1950–1959; doi:10.1097/HJH.0000000000003198.

Tan X, Chen G, Chen K. 2023. Clean heating and air pollution: Evidence from Northern China. Energy Reports 9:303–313; doi:10.1016/j.egyr.2022.11.166.

Thomas E, Wickramasinghe K, Mendis S, Roberts N, Foster C. 2015. Improved stove interventions to reduce household air pollution in low and middle income countries: A descriptive systematic review. BMC Public Health 15:650; doi:10.1186/s12889-015-2024-7.

Thompson RJ, Li J, Weyant CL, Edwards R, Lan Q, Rothman N, et al. 2019. Field Emission Measurements of Solid Fuel Stoves in Yunnan, China Demonstrate Dominant Causes of Uncertainty in Household Emission Inventories. Environmental Science & Technology 53:3323–3330; doi:10.1021/acs.est.8b07040.

Tuck MK, Chan DW, Chia D, Godwin AK, Grizzle WE, Krueger KE, et al. 2009. Standard

Operating Procedures for Serum and Plasma Collection: Early Detection Research Network Consensus Statement *Standard Operating Procedure Integration Working Group*. Journal of Proteome Research 8:113–117; doi:10.1021/pr800545q.

van Buuren S, Groothuis-Oudshoorn K. 2011. **Mice** : Multivariate Imputation by Chained Equations in *R*. Journal of Statistical Software 45; doi:10.18637/jss.v045.i03.

VanderWeele TJ. 2015. *Explanation in causal inference: Methods for mediation and interaction*. Oxford University Press:New York.

Volckens J, Quinn C, Leith D, Mehaffy J, Henry CS, Miller-Lionberg D. 2017. Development and evaluation of an ultrasonic personal aerosol sampler. Indoor air 27:409–416; doi:10.1111/ina.12318.

Wang Q, Zhao Q, Wang G, Wang B, Zhang Y, Zhang J, et al. 2020. The association between ambient temperature and clinical visits for inflammation-related diseases in rural areas in China. Environmental Pollution 261:114128; doi:10.1016/j.envpol.2020.114128.

Wen H, Nie P, Liu M, Peng R, Guo T, Wang C, et al. 2023. Multi-health effects of clean residential heating: Evidences from rural China's coal-to-gas/electricity project. Energy for Sustainable Development 73:66–75; doi:10.1016/j.esd.2023.01.013.

Wooldridge JM. 2021. Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators.; doi:10.2139/ssrn.3906345.

Xu H, Brook RD, Wang T, Song X, Feng B, Yi T, et al. 2019. Short-term effects of ambient air pollution and outdoor temperature on biomarkers of myocardial damage, inflammation and oxidative stress in healthy adults. Environmental Epidemiology 3:e078; doi:10.1097/EE9.0000000000000078.

Yan L, Carter E, Fu Y, Guo D, Huang P, Xie G, et al. 2020. Study protocol: The INTERMAP China Prospective (ICP) study. Wellcome Open Research 4:154; doi:10.12688/wellcomeopenres.15470.2.

Yang K. 2021. Power industry provides inexhaustible power for national rejuvenation; .

Yap P-S, Garcia C. 2015. Effectiveness of Residential Wood-Burning Regulation on Decreasing Particulate Matter Levels and Hospitalizations in the San Joaquin Valley Air Basin. American Journal of Public Health 105:772–778; doi:10.2105/AJPH.2014.302360.

Ye W, Steenland K, Quinn A, Liao J, Balakrishnan K, Rosa G, et al. 2022. Effects of a Liquefied Petroleum Gas Stove Intervention on Gestational Blood Pressure: Intention-to-Treat and Exposure-Response Findings From the HAPIN Trial. Hypertension 79:1887–1898; doi:10.1161/HYPERTENSIONAHA.122.19362.

Young OR, Guttman D, Qi Y, Bachus K, Belis D, Cheng H, et al. 2015. Institutionalized governance processes: Comparing environmental problem solving in China and the United States. Global Environmental Change 31:163–173; doi:10.1016/j.gloenvcha.2015.01.010.

Yu C, Kang J, Teng J, Long H, Fu Y. 2021. Does coal-to-gas policy reduce air pollution? Evidence from a quasi-natural experiment in China. Science of The Total Environment 773:144645; doi:10.1016/j.scitotenv.2020.144645.

Yun X, Shen G, Shen H, Meng W, Chen Y, Xu H, et al. 2020. Residential solid fuel emissions contribute significantly to air pollution and associated health impacts in China. Science Advances

6:eaba7621; doi:10.1126/sciadv.aba7621.

Zhang J(Jim), Smith KR. 2007. Household Air Pollution from Coal and Biomass Fuels in China: Measurements, Health Impacts, and Interventions. Environmental Health Perspectives 115:848–855; doi:10.1289/ehp.9479.

Zhuang Z, Li Y, Chen B, Guo J. 2009. Chinese kang as a domestic heating system in rural northern China—A review. Energy and Buildings 41:111–119; doi:10.1016/j.enbuild.2008.07.013.

Zigler CM, Kim C, Choirat C, Hansen JB, Wang Y, Hund L, et al. 2016. *Causal inference methods for estimating long-term health effects of air quality regulations. Research report 187.* Health Effects Institute / Health Effects Institute:Boston, MA.

A Appendices

A.1 Biomarker descriptives

Below we show boxplots for the logged values of the blood inflammatory and oxidative stress markers.

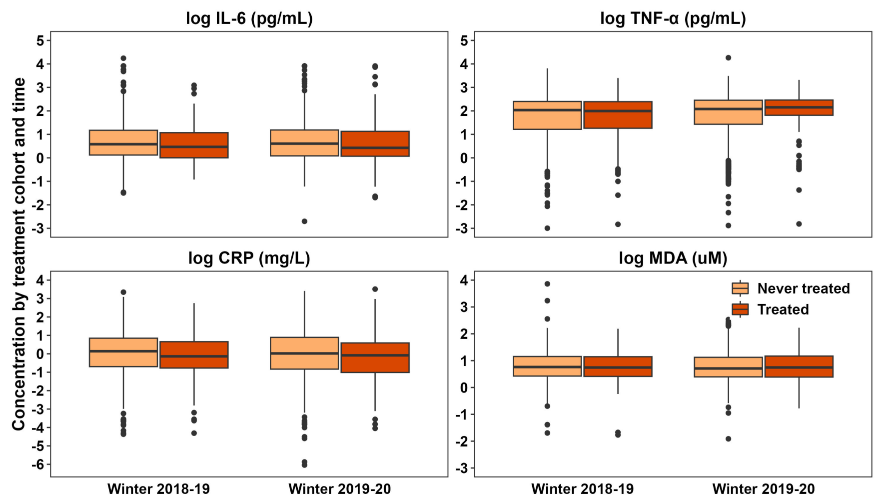
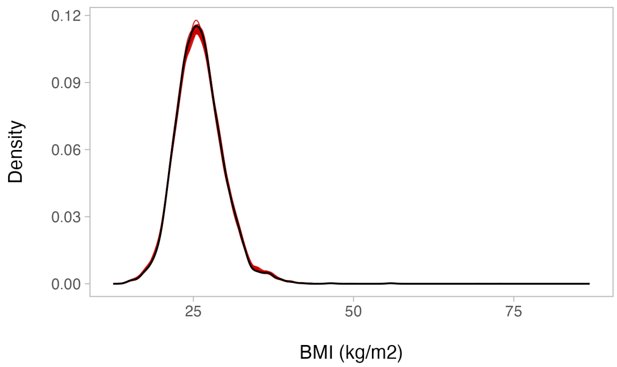


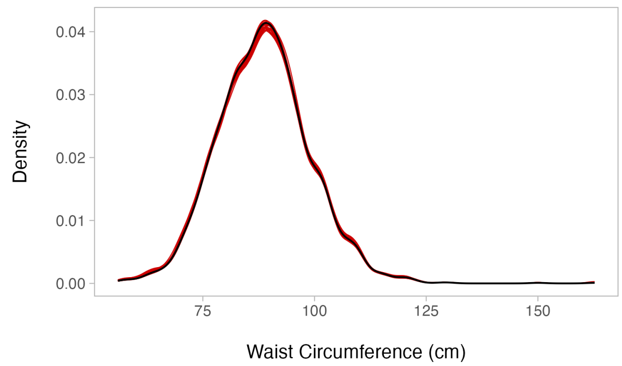
Figure A1: Boxplots for markers of systemic inflammation including C-reactive protein (CRP), interleukin-6 (IL-6), tumour necrosis factor alpha (TNF-α) and markers of oxidative stress including 8-hydroxy-2'-deoxyguanosine (8-OHdG) and malondialdehyde (MDA)

A.2 Imputation results

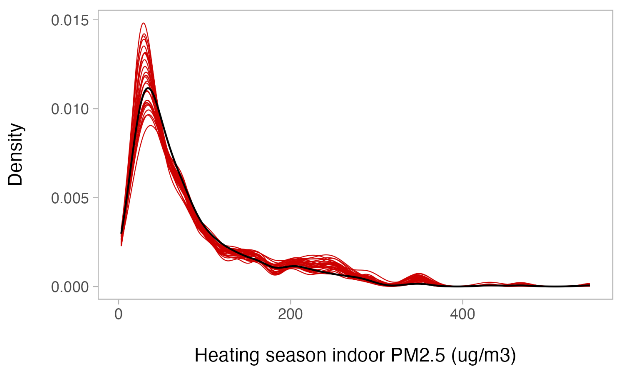
The figures below show density plots for the values of body mass index, waist circumference, and indoor $PM_{2.5}$ from the multiple imputation models. The red lines show the values for each of the 30 imputed datasets, and the black line shows the value for the observed data.



(a)



(b)



(c)

Figure A2: Kernal density plots showing distribution of multiply-imputed values for body mass index (kg/m^2), waist circumference (cm), and indoor $PM_{2.5}$ ($\mu g/m^3$) (red lines) and observed values (heavy black line)

A.3 Participant flow diagram

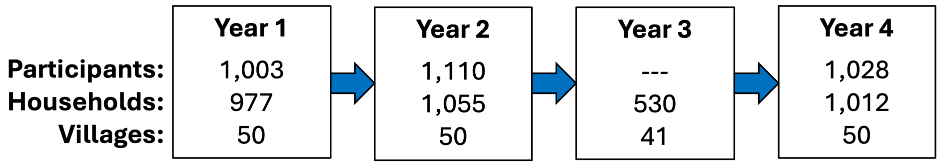


Figure A3: Flow chart of BHET study participation at the participant, household, and village levels across study years.

A.4 Policy uptake

Figure A4 shows trends over time in self-reported coal and biomass consumption over each season. Table A1 shows results from applying our extended two-way fixed effects models (in separate analyses) to coal and biomass consumption.

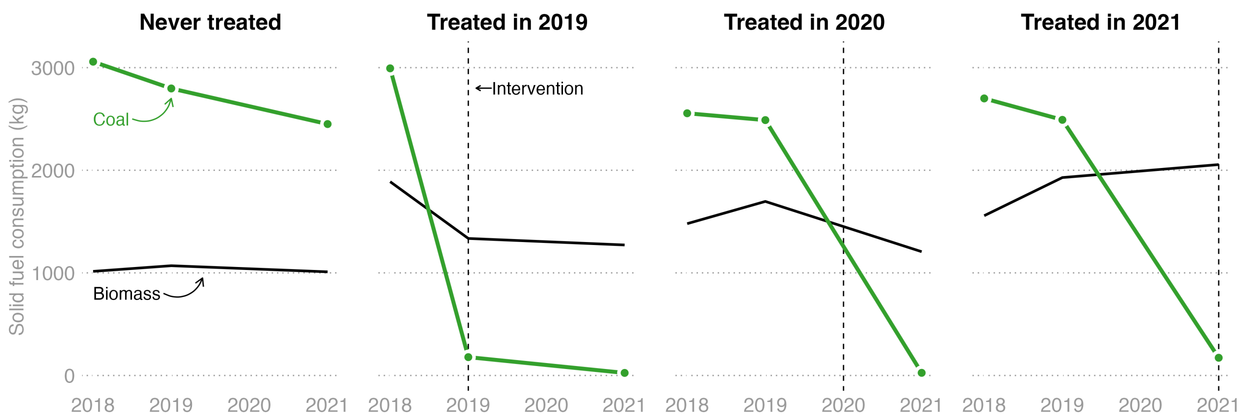


Figure A4: Trends in self-reported coal and biomass, by treatment season

Table A1: Policy impacts on self-reported fuel use (kg)

		Coal[a]		Biomass[b]	
Cohort	Time	ATT	(95%CI)	ATT	(95%CI)
Average ATT					
All	All	-2361	(-2677, -2044)	-487	(-805, -168)
Cohort-Time ATTs					
2019	2019	-2631	(-2913, -2348)	-653	(-991, -315)
2019	2021	-2416	(-2847, -1984)	-633	(-1201, -64)
2020	2021	-2018	(-2474, -1562)	-350	(-701, 0)
2021	2021	-1961	(-2895, -1027)	338	(-30, 705)

[a] Joint test that all ATTs are equal: $F(3, 2886) = 1.856$, $p = 0.135$

[b] Joint test that all ATTs are equal: $F(3, 2886) = 5.545$, $p = 0.001$

A.5 Heterogeneity in treament effects

A.5.1 Personal exposure

As noted in the methods section...Table Table A2 shows limited evidence that the ATTs across cohorts and time demonstrate meaningful heterogeneity.

Table A2: Heterogenous treatment effects: Personal exposures

		PM2.5[a]		Black carbon[b]	
Cohort	Time	ATT	(95%CI)	ATT	(95%CI)
Average ATT					
All	All	1.95	(-23.34, 27.23)	-0.43	(-1.67, 0.81)
Cohort-Time ATTs					
2019	2019	-0.05	(-28.97, 28.87)	-0.69	(-1.84, 0.45)
2019	2021	-4.31	(-41.92, 33.3)	-0.25	(-2.11, 1.62)
2020	2021	23.61	(-19.88, 67.11)	-0.27	(-2.04, 1.5)
2021	2021	-19.06	(-43.19, 5.07)	-0.56	(-2.46, 1.34)

[a] Joint test that all ATTs are equal: $F(3, 1271)= 0.431$, $p= 0.731$

[b] Joint test that all ATTs are equal: $F(3, 1253)= 0.613$, $p= 0.607$

A.5.2 Indoor $PM_{2.5}$

Table Table A3 shows estimates for cohort-time ATTs for daily and seasonal indoor $PM_{2.5}$.

Table A3: Heterogenous treatment effects: Indoor

		Daily[a]		Seasonal[b]	
Cohort	Time	ATT	(95%CI)	ATT	(95%CI)
Average ATT					
All	All	-14.20	(-53.94, 25.54)	-36.19	(-60.74, -11.65)
Cohort-Time ATTs					
2020	2021	-4.71	(-56.93, 47.5)	-25.44	(-58.02, 7.13)
2021	2021	-37.24	(-74.15, -0.33)	-59.23	(-79.61, -38.85)

[a] Joint test that all ATTs are equal: $F(1, 405)= 0.064$, $p= 0.8$

[b] Joint test that all ATTs are equal: $F(1, 368)= 0.756$, $p= 0.385$

A.5.3 Indoor temperature

Table A4: Heterogenous treatment effects: Indoor temperature

		Point temp (°C)			Mean temp (°C)			Min temp (°C)		
Cohort	Time	ATT	(95%CI)	p-value	ATT	(95%CI)	p-value	ATT	(95%CI)	p-value
All times										
2019	2019	1.77	(0.66, 2.88)		0.43	(-0.71, 1.57)		1.96	(0.43, 3.48)	
	2020				0.52	(-0.22, 1.26)		2.42	(0.54, 4.3)	
	2021	2.29	(0.51, 4.07)		0.79	(0, 1.57)		4.93	(2.28, 7.58)	
2020	2020				0.87	(-0.2, 1.93)		5.00	(3.22, 6.79)	
	2021	2.36	(0.54, 4.17)		0.58	(-0.66, 1.82)		6.87	(4.35, 9.39)	
2021	2021	0.64	(-1.08, 2.35)	0.440	1.06	(0.32, 1.79)	0.320	2.04	(0.08, 4)	0.000
Daytime										
2019	2019				0.44	(-0.96, 1.83)				
	2020				1.26	(0.36, 2.17)				
	2021				1.50	(0.55, 2.46)				
2020	2020				0.28	(-1.45, 2.02)				
	2021				0.13	(-1.7, 1.97)				
2021	2021				1.44	(0.64, 2.25)	0.260			
Daytime heating										
2019	2019				0.80	(-0.48, 2.09)				
	2020				1.43	(0.04, 2.83)				
	2021				2.33	(1.03, 3.62)				
2020	2020				2.63	(1.87, 3.39)				
	2021				2.46	(1.46, 3.46)				
2021	2021				2.13	(0.67, 3.59)	0.000			
Heating season										
2019	2019				1.05	(-0.1, 2.2)		1.94	(0.42, 3.47)	
	2020				1.23	(-0.11, 2.58)		2.41	(0.53, 4.3)	
	2021				2.07	(0.88, 3.27)		5.34	(2.66, 8.02)	
2020	2020				2.71	(2.04, 3.37)		4.35	(3.17, 5.53)	
	2021				2.48	(1.33, 3.62)		6.27	(3.73, 8.81)	
2021	2021				1.97	(0.53, 3.41)	0.000	2.23	(0.26, 4.21)	0.000

A.5.4 Blood pressure outcomes

Table A5 shows ATTs by treatment cohort and time, as well as the results of joint tests of heterogeneity across ATTs.

Table A5: Heterogenous treatment effects for the total effect of the CBHP policy on blood pressure.

		Adjusted DiD[a]		Heterogeneity tests[b]	
Cohort	Time	ATT	(95%CI)	F-Statistic	p-value
Brachial SBP					
2019	2019	-2.36	(-5.23, 0.5)		
2019	2021	-1.51	(-4.01, 0.98)		
2020	2021	-1.26	(-4.97, 2.45)		
2021	2021	2.39	(-0.49, 5.28)	2.3	0.080
Central SBP					
2019	2019	-2.03	(-4.69, 0.63)		
2019	2021	-1.96	(-4.45, 0.52)		
2020	2021	-1.78	(-5.07, 1.52)		
2021	2021	2.11	(-1.09, 5.31)	1.9	0.140
Brachial DBP					
2019	2019	-2.66	(-4.67, -0.65)		
2019	2021	-2.37	(-4.01, -0.72)		
2020	2021	0.2	(-1.54, 1.94)		
2021	2021	0.78	(-0.48, 2.05)	6.8	0.000
Central DBP					
2019	2019	-2.67	(-4.57, -0.78)		
2019	2021	-2.55	(-4.15, -0.94)		
2020	2021	0.11	(-1.67, 1.9)		
2021	2021	1.09	(-0.06, 2.23)	10.0	0.000

Note: ATT = Average Treatment Effect on the Treated, DiD = Difference-in-Differences, CDE = Controlled Direct Effect.

[a] Adjusted for age, sex, waist circumference, smoking, alcohol consumption, and use of blood pressure medication.

[b] F-statistics and p-values for joint tests of equality across cohort and time ATTs

A.5.5 Mediation analyses for blood pressure

Table A6 shows the cohort-time treatment effects for the mediation model for blood pressure.

Table A6: Heterogenous treatment effects for blood pressure mediation model

		Adjusted Total Effect[a]		CDE Mediated By:[b]					
				Indoor PM		Indoor Temp		PM + Temp	
Cohort	Time	ATT	(95%CI)	ATT	(95%CI)	ATT	(95%CI)	ATT	(95%CI)
Brachial SBP									
2019	2019	-2.36	(-5.23, 0.50)	-2.15	(-5.14, 0.84)	-1.69	(-4.54, 1.15)	-1.24	(-4.20, 1.72)
2019	2021	-1.51	(-4.01, 0.98)	-1.27	(-4.01, 1.47)	-0.41	(-2.92, 2.10)	0.01	(-2.71, 2.74)
2020	2021	-1.26	(-4.97, 2.45)	-0.54	(-4.25, 3.17)	0.43	(-2.86, 3.73)	1.04	(-2.59, 4.67)
2021	2021	2.39	(-0.49, 5.28)	2.68	(-0.42, 5.79)	1.95	(-1.74, 5.64)	1.88	(-1.92, 5.67)
Central SBP									
2019	2019	-2.03	(-4.69, 0.63)	-1.75	(-4.61, 1.11)	-1.40	(-4.06, 1.27)	-0.89	(-3.73, 1.95)
2019	2021	-1.96	(-4.45, 0.52)	-1.65	(-4.40, 1.11)	-0.93	(-3.18, 1.32)	-0.44	(-2.95, 2.07)
2020	2021	-1.78	(-5.07, 1.52)	-1.00	(-4.36, 2.36)	-0.15	(-3.18, 2.88)	0.47	(-2.95, 3.89)
2021	2021	2.11	(-1.09, 5.31)	2.45	(-0.83, 5.73)	1.66	(-1.73, 5.05)	1.63	(-1.82, 5.08)
Brachial DBP									
2019	2019	-2.66	(-4.67, -0.65)	-2.47	(-4.70, -0.25)	-2.29	(-4.18, -0.40)	-1.94	(-4.03, 0.14)
2019	2021	-2.37	(-4.01, -0.72)	-2.10	(-4.09, -0.11)	-1.81	(-3.21, -0.41)	-1.50	(-3.28, 0.27)
2020	2021	0.20	(-1.54, 1.94)	0.31	(-1.43, 2.04)	1.14	(-0.65, 2.94)	1.23	(-0.70, 3.15)
2021	2021	0.78	(-0.48, 2.05)	1.05	(-0.59, 2.69)	0.20	(-1.21, 1.62)	0.36	(-1.34, 2.06)
Central DBP									
2019	2019	-2.67	(-4.57, -0.78)	-2.43	(-4.58, -0.28)	-2.52	(-4.34, -0.70)	-2.13	(-4.18, -0.08)
2019	2021	-2.55	(-4.15, -0.94)	-2.20	(-4.18, -0.22)	-2.18	(-3.60, -0.76)	-1.80	(-3.58, -0.03)
2020	2021	0.11	(-1.67, 1.90)	0.22	(-1.58, 2.01)	1.07	(-0.74, 2.87)	1.16	(-0.80, 3.13)
2021	2021	1.09	(-0.06, 2.23)	1.39	(-0.16, 2.94)	0.51	(-0.80, 1.82)	0.70	(-0.94, 2.34)

Note: Results combined across 30 multiply-imputed datasets. ATT = Average Treatment Effect on the Treated, CDE = Controlled Direct Effect, DBP = Diastolic blood pressure, SBP = Systolic blood pressure.

[a] Adjusted for age, sex, waist circumference, smoking, alcohol consumption, and use of blood pressure medication.

[b] Mediators were set to the mean value for untreated participants at baseline.

Table A7: Heterogenous treatment effects and tests for cohort-time heterogeneity across CDEs for multiple mediation blood pressure mediation model.

		Adjusted CDE[a]		Heterogeneity tests[b]	
Cohort	Time	ATT	(95%CI)	F-Statistic	p-value
Brachial SBP					
2019	2019	-1.24	(-4.20, 1.72)		
2019	2021	0.01	(-2.71, 2.74)		
2020	2021	1.04	(-2.59, 4.67)		
2021	2021	1.88	(-1.92, 5.67)	0.8	0.513
Central SBP					
2019	2019	-0.89	(-3.73, 1.95)		
2019	2021	-0.44	(-2.95, 2.07)		
2020	2021	0.47	(-2.95, 3.89)		
2021	2021	1.63	(-1.82, 5.08)	0.6	0.608
Brachial DBP					
2019	2019	-1.94	(-4.03, 0.14)		
2019	2021	-1.50	(-3.28, 0.27)		
2020	2021	1.23	(-0.70, 3.15)		
2021	2021	0.36	(-1.34, 2.06)	3.9	0.008
Central DBP					
2019	2019	-2.13	(-4.18, -0.08)		
2019	2021	-1.80	(-3.58, -0.03)		
2020	2021	1.16	(-0.80, 3.13)		
2021	2021	0.70	(-0.94, 2.34)	4.8	0.003

Note: ATT = Average Treatment Effect on the Treated, DiD = Difference-in-Differences, CDE = Controlled Direct Effect.

[a] Adjusted for age, sex, waist circumference, smoking, alcohol consumption, and use of blood pressure medication.

[b] F-statistics and p-values for joint tests of equality across cohort and time ATTs

A.5.6 Respiratory outcomes

Appendix tables A8, A9, A10, A11, A12, A13 below show Average Treatment Effect on the Treated (ATTs) by treatment cohort and time. ATTs are derived from estimating marginal effects from extended two-way fixed effects models with additional adjustment for age, sex, and smoking status.

Table A8: Heterogenous treatment effects for self-reported respiratory outcomes: Any respiratory symptom

Cohort	Time	ATT	(95%CI)
Average ATT			
All	All	-0.08	(-0.15, -0.01)
Cohort-Time ATTs			
2019	2019	-0.11	(-0.20, -0.02)
2019	2021	-0.10	(-0.21, 0.00)
2020	2021	0.01	(-0.10, 0.13)
2021	2021	-0.12	(-0.22, -0.01)

Note: Joint test that all ATTs are equal: F(3, 2579)= 1.283, p= 0.278.

Table A9: Heterogenous treatment effects for self-reported respiratory outcomes: Coughing

Cohort	Time	ATT	(95%CI)
Average ATT			
All	All	-0.02	(-0.07, 0.03)
Cohort-Time ATTs			
2019	2019	-0.04	(-0.11, 0.03)
2019	2021	0.01	(-0.07, 0.08)
2020	2021	-0.03	(-0.10, 0.05)
2021	2021	-0.04	(-0.09, 0.02)

Note: Joint test that all ATTs are equal: F(3, 2579)= 0.732, p= 0.533.

Table A10: Heterogenous treatment effects for self-reported respiratory outcomes: Phlegm

Cohort	Time	ATT	(95%CI)
Average ATT			
All	All	-0.02	(-0.06, 0.03)
Cohort-Time ATTs			
2019	2019	-0.06	(-0.16, 0.03)
2019	2021	-0.03	(-0.10, 0.04)
2020	2021	0.04	(-0.02, 0.09)
2021	2021	0.03	(-0.04, 0.09)

Note: Joint test that all ATTs are equal: $F(3, 2579) = 1.735$, $p = 0.158$.

Table A11: Heterogenous treatment effects for self-reported respiratory outcomes: Wheezing attacks

Cohort	Time	ATT	(95%CI)
Average ATT			
All	All	0.00	(-0.04, 0.04)
Cohort-Time ATTs			
2019	2019	-0.02	(-0.06, 0.01)
2019	2021	0.01	(-0.04, 0.06)
2020	2021	-0.03	(-0.11, 0.05)
2021	2021	0.09	(-0.00, 0.18)

Note: Joint test that all ATTs are equal: $F(3, 2579) = 2.923$, $p = 0.033$.

Table A12: Heterogenous treatment effects for self-reported respiratory outcomes: Trouble breathing

Cohort	Time	ATT	(95%CI)
Average ATT			
All	All	-0.05	(-0.12, 0.02)
Cohort-Time ATTs			
2019	2019	-0.06	(-0.16, 0.04)
2019	2021	-0.07	(-0.16, 0.03)
2020	2021	0.01	(-0.09, 0.11)
2021	2021	-0.07	(-0.20, 0.06)

Note: Joint test that all ATTs are equal: $F(3, 2579) = 0.718$, $p = 0.541$.

Table A13: Heterogenous treatment effects for self-reported respiratory outcomes: Chest trouble

Cohort	Time	ATT	(95%CI)
Average ATT			
All	All	-0.06	(-0.12, -0.01)
Cohort-Time ATTs			
2019	2019	-0.06	(-0.13, 0.01)
2019	2021	-0.06	(-0.15, 0.03)
2020	2021	-0.05	(-0.16, 0.05)
2021	2021	-0.14	(-0.22, -0.05)

Note: Joint test that all ATTs are equal: F(3, 2579)= 1.046, p= 0.371.

A.5.7 Outdoor and personal mixed combustion

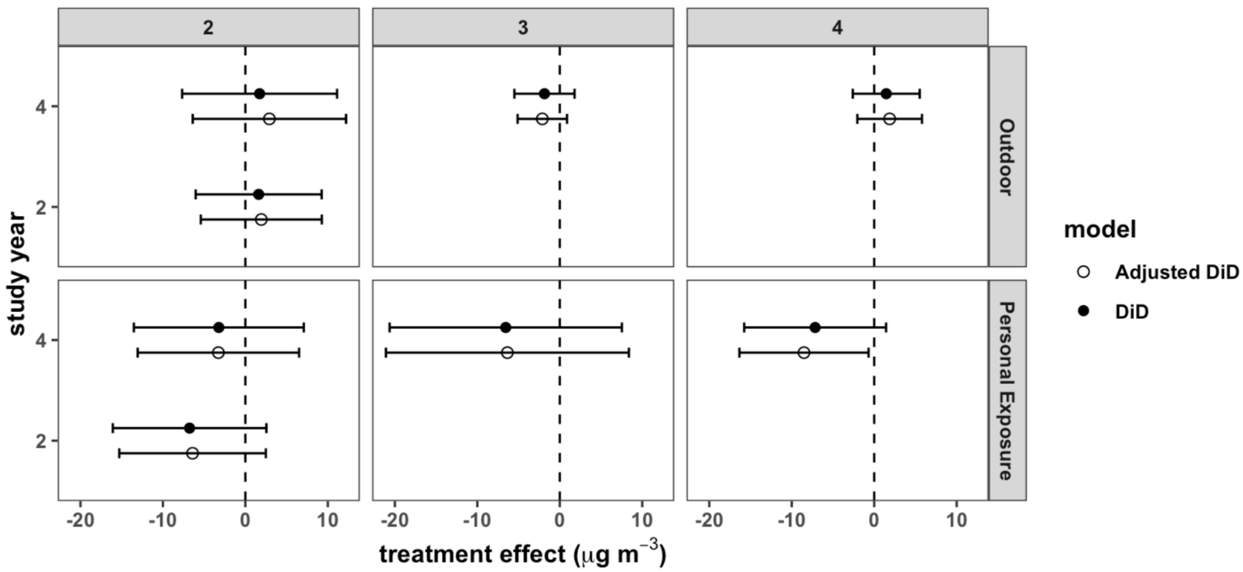


Figure A5: Adjusted and unadjusted treatment effect for outdoor and personal exposure (μg/m^3) to the mixed combustion source by treatment year.

A.6 Impact of sample composition on FeNO results

Table A14 shows differences in the ATTs for the impact of the CBHP policy on FeNO depending on whether the estimation sample includes all individuals or is limited to those with repeated measures across campaigns.

Table A14: Effects of the CBHP policy on FeNO (ppb) based on the number of individuals with repeated measurements.

	All participants		Participants with >1 measure		Participants with 3 measures	
	ATT	(95%CI)	ATT	(95%CI)	ATT	(95%CI)
DiD	0.17	(-2.24, 2.58)	0.13	(-3.09, 3.35)	-0.57	(-3.08, 1.94)
Adjusted DiD	0.55	(-2.03, 3.13)	0.24	(-3.19, 3.67)	0.27	(-2.39, 2.92)
Observations	794		526		252	

A.7 Impact of including Season 3 data

Table A15 shows differences in the ATTs for the impact of seasonal indoor $PM_{2.5}$ when season 3 data (collected in 41 villages during COVID-19) are included versus excluded.

Table A15: Effects of the CBHP policy on indoor seasonal $PM_{2.5}$ based on whether Season 3 data are included vs. excluded.

		With Season 3 data		Without Season 3 data	
Cohort	Time	ATT	(95%CI)	ATT	(95%CI)
Average ATT					
All	All	-37.49	(-60.11, -14.88)	-35.11	(-59.36, -10.85)
Cohort-Time ATTs					
2020	2020	-36.94	(-61.39, -12.49)	0.00	(NA, NA)
2020	2021	-33.51	(-66.84, -0.18)	-30.22	(-63.77, 3.32)
2021	2021	-46.82	(-58.57, -35.07)	-44.88	(-60.41, -29.34)

Note:

Sample sizes for

About the authors

Other publications

Li X, Baumgartner J, Barrington-Leigh C, Harper S, Robinson B, Shen G, et al. 2022a. Socioeconomic and Demographic Associations with Wintertime Air Pollution Exposures at Household, Community, and District Scales in Rural Beijing, China. Environ Sci Technol 56:8308–8318; doi:10.1021/acs.est.1c07402.

Li X, Baumgartner J, Harper S, Zhang X, Sternbach T, Barrington-Leigh C, et al. 2022b. Field measurements of indoor and community air quality in rural Beijing before, during, and after the COVID-19 lockdown. Indoor Air 32:e13095; doi:10.1111/ina.13095.

Sternbach TJ, Harper S, Li X, Zhang X, Carter E, Zhang Y, et al. 2022. Effects of indoor and outdoor temperatures on blood pressure and central hemodynamics in a wintertime longitudinal study of Chinese adults. J Hypertension 40:1950–1959; doi:10.1097/HJH.0000000000003198.

[a](**ellison.carter?**)(**gmail.com?**) Do these numbers show air pollution levels in S2? While, these numbers do not match with S2 in Table 5. *Assigned to ellison.carter@gmail.com* [b](**xiaoyingcsu?**)(**gmail.com?**) Can you help me verify that this is true? Based on the reductions observed, I think indoor PM2.5 mass concentrations in treated homes should have reached IT-4, but I'm not sure. *Assigned to xiaoyingcsu@gmail.com* [c](**talia.sternbach?**)(**gmail.com?**). DO NOT READ THIS UNTIL AFTER THE WEEKEND!! But if you can drop in the figures you created awhile back, that would be great. [d]This needs organization and punchier writing but I'm losing steam. [e](**talia.sternbach?**)(**gmail.com?**). To discuss in the meeting on Monday.