DISSERTATION

NOVEL APPLICATIONS OF DATA-DRIVEN APPROACHES FOR UNDERSTANDING THE IMPACTS OF HOUSEHOLD ENERGY INTERVENTIONS

Submitted by

Collin Brehmer

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2025

Doctoral Committee:

Advisor: Ellison Carter

Frances Davenport
Kayleigh Keller
Sybil Sharvelle

ABSTRACT

NOVEL APPLICATIONS OF DATA-DRIVEN APPROACHES FOR UNDERSTANDING THE IMPACTS OF HOUSEHOLD ENERGY INTERVENTIONS

Air pollution from household solid fuel combustion is associated with premature death, disease, and radiative climate forcing.[1,2] Beginning in 2015, the Chinese government implemented the Clean Heating Policy in Northern China (CHP) with the goal to transition 70% of homes in the Beijing region from coal-based space heating to natural gas or electric-powered space heating. Studies of the impact of the CHP on air pollution and the potential mechanisms of action are limited.

The continued use of a secondary solid fuel or heating device after the primary solid fuel heating stove is replaced with a cleaner alternative could weaken the impacts of the effort to replace the primary solid fuel stove. In Chapter 1, we identified heating events from biomass kang stoves as a proxy for stove use using a combination of manually labeled data and XGBoost modeling. We showed that biomass kang stove usage did not change because of the CHP and agreed with self-reported measures of heating duration. Our results demonstrated the capability of XGBoost to identify stove use events when trained on manually labeled event data and provided evidence that self-reported measures of stove use may be sufficient for understanding how secondary stove use changes as a result of a household energy intervention.

Fine particulate matter air pollution ($PM_{2.5}$) is of particular interest when evaluating household energy transitions since it is a product of incomplete combustion and is related to several health outcomes. We evaluated the impacts of the CHP on seasonal indoor, seasonal

outdoor, and 24-hr personal $PM_{2.5}$ exposure in 50 villages, 300 homes, and 500 participants during three years over a four-year period. The CHP had high uptake, with a significant decrease in coal usage in treated groups. We also observed a significant reduction in seasonal average indoor $PM_{2.5}$ (22.2 [4.2, 40.3] $\mu g/m^3$). Seasonal outdoor and 24-hr personal $PM_{2.5}$ exposure did decrease over time but the decrease could not be attributed to the CHP due to similar decreases in treated and untreated groups. Our study suggests that the CHP yielded promising results in reducing indoor $PM_{2.5}$ and provided valuable insights for household energy transitions worldwide.

Given that most household energy interventions target one source of air pollution, using a mixture of sources as an outcome, like $PM_{2.5}$, when only one of the sources of air pollution is targeted by the policy can make it hard to disentangle the effects of the policy if the variability in the non-targeted sources is high. Chapter 3 identified sources and their contributions to outdoor and personal $PM_{2.5}$ exposure using chemical analysis and source apportionment. We used the concentration of the coal-containing source in outdoor and personal exposure measurements as the outcome in policy analysis models and compared the findings to the models where total $PM_{2.5}$ was the outcome. We found a significant reduction in personal exposure to the coal containing source (-7.75 [-13.4, -2.14] $\mu g\ m^{-3}$), which contrasts with our findings that the CHP had no impact on personal exposure to total $PM_{2.5}$. This work demonstrates how additional granularity in the air pollution outcome can serve as a better outcome than a mixture of sources.

ACKNOWLEDGEMENTS

Before thanking anyone in my graduate career, I first want to thank Dr. Jamie Schauer, Dr. Alex Lai, and Dr. Ross Edwards at UW-Madison. Through Jamie's lab, I was able to gain experience with the type of air pollution work we do, and in work my passion for air quality research was ignited. I am forever grateful to Jamie for taking a chance on a junior undergrad to lead and first author three papers. To Alex, I am so thankful that you were the PhD trainee that I worked under. You helped foster my curiosity and answered every question I had while still letting me have autonomy. To Ross, I am so appreciative of your interest in the physical science side and helping me step outside of my comfort zone to work more on the applied engineering. I learned so much spending time hacking an inkjet printer, soldering the boards, coding them, machining the for the printer, and using your 3-d printer. The work I did with you three couldn't have set me up better to have a successful graduate career.

The most important person to thank for my graduate school journey is my advisor, Dr. Ellison Carter. Ellison and I started working together in 2017 when I was hired as an undergrad student to, like most of us in this field started out, organize thousands of air pollution filter samples for a project she was leading. Thank you for helping me struggle through writing my first paper and putting up with the several drafts, without this help I probably would not have felt as passionate about the type of air pollution research we are continuing to this day, close to 8 years later. I have appreciated your support and encouragement to explore areas of research outside of the typical engineering box. I am a more thoughtful, intentional, and confident scientist because of you, for which I will be forever grateful.

I have to acknowledge the BHET team. I don't think that I could have pictured a better team of scientist to work with. Particularly, I want to thank Dr. Xiaoying Li for her collaboration leading and processing the air pollution data, and co- first authoring the air pollution policy paper with me. I also must thank Talia Sternbach, Kennedy Hirst, and Enkhuun Byambadorj for spending dozens of hours in the summer looking at stove temperature plots. I have learned so much from everyone on the team and would not be half the scientist I am today without the team.

To all my friends outside of academia who kept me sane and recognized me as a person first and a scientist second. Cameron, Mike, Dillan, Joe, Alec, Kyle, and Joshua, without all of you graduate school would have been a lonely, isolating, experience focused only on work. Coming into graduate school I put too much of my self-worth on the work that I was doing, and all of you helped me realize that I was more than just the work that I produced. You all shifted my view on life and helped me be a happier person.

This work is dedicated to those that are irreplicable. To my parents, who supported me in all aspects of the word for entire life. There was never a doubt in my mind that I could call and ask them for anything, within reason, and I would be supported in my goals to get a PhD. To my partner Jason, who has stuck with me for the entire six years. I am so grateful that you upended your life to move to Fort Collins during the pandemic and returned to Denver after you finished your grad school. You make me a better person and I could not have done this without you.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION AND BACKGROUND

1.1 Health and environmental impacts of air pollution

Approximately one third (~2.1 billion people) of the global population burns solid fuel sources like coal or biomass to cook or heat their homes.[3] Solid fuels are combusted in a wide variety of stoves that vary in their combustion efficiency and subsequent air pollution emissions. Generally, solid fuel stoves are inefficient, lack appropriate control mechanisms like ventilation, and are major sources of indoor, outdoor, and personal air pollution exposure in homes that use them.

Homes that use solid fuel for cooking and/or heating have air pollution concentrations that are considerably higher than any recommended guidelines.[4] Solid fuel stoves tend to be used the most in rural, low-income areas that have additional risk factors for premature death and disease. Air pollution in general has been linked to a wide variety of diseases including obesity, diabetes, neurological disease, and cardiovascular disease.[5–7] Specifically, exposure to air pollution from solid fuel combustion is a leading environmental risk factor for premature death and was estimated to contribute to 3.2 million deaths in 2020 alone.[3] The best understood mechanism of disease is that the small inhaled particles infiltrate deep into the lungs, enter the bloodstream, and can cause systematic inflammation which has been associated with several chronic diseases.[8,9]

Solid fuel combustion also emits multiple pollutants that are relevant to climate. Particle phase pollutants like black carbon are light absorbing and have been found to contribute to radiative forcing.[1] Both particle and gas phase pollutants can impact the composition of the atmosphere by changing the properties of clouds and generating secondary organic and inorganic aerosols. Gas phase volatile organic compounds can react with nitrogen compounds to contribute to ground level ozone formation.[10] Other gas phase pollutants like methane and carbon dioxide directly contribute to global environmental change through the greenhouse effect.[11,12]

1.2 Air pollution policies in China

After severe haze events in China during the winter of 2013, the focus of air pollution control measures shifted towards reducing $PM_{2.5}$ levels across China through the "war against pollution".[13,14] Ten tasks were initiated due to the haze events that had the general goals of improving air quality through increasing the clean energy supply, introducing additional air pollution control measures, establishing monitoring networks, and creating regulatory bodies. These ten tasks sparked the modern push for the stricter $PM_{2.5}$ regulations and control policies being implemented currently.

Northern China experiences cold, dry winters that require heating to maintain comfortable indoor temperatures. Central heating systems are uncommon in rural areas of China and most homes use coal and/or biomass kang stoves for heating in the winter. Solid fuel combustion in rural China contributed up to an estimated 21% of outdoor air pollution in the wintertime, meaning household solid fuel combustion is an issue beyond indoor air quality.[15] Estimates attribute 366,000 premature deaths annually in China due to air pollution from coal combustion.[16]

Recognizing the impacts of household solid fuel combustion on air quality and health, the Chinese government implemented the Clean Heating Program for Northern China (CHP) in 2016 with the goal of achieving 70% clean household heating by 2021. Before the implementation of the CHP, 80% of the rural residential heating energy needs for over 100 million homes was met by burning roughly 200 million tons of coal.[17] The CHP took advantage of the near 100% electrification of homes in 2016 and provided subsides for the purchase and installation of natural gas or electric powered heat pumps as well as subsides for nighttime electricity rates.

1.3 Previous evaluations of household energy transitions

Household energy interventions have mostly focused on the replacement of solid fuel based cookstoves with more efficient stoves that use cleaner fuels. The implementation of cleaner cookstoves should reduce air pollution and improve health outcomes, but evidence from real-world intervention evaluations have shown mixed findings.[18] Most previous work has evaluated interventions administered at a small scale in a semi-experimental design. Observational studies of real-world household energy transitions have rarely been studied and remain further under evaluated at the local level.

Some evaluations of larger scale clean energy programs have been conducted in high-income countries. Policies in Ireland (residential coal ban), Australia (wood stove exchange), and New Zealand (general clean energy policies) improved black smoke or PM_{10} concentrations.[19–21] Several modeling studies have been conducted to understand the potential impacts of the CHP on outdoor and indoor $PM_{2.5}$, finding significant reductions in air pollution. One modeling study estimated a 36% reduction in personal exposure attributed to fuel use changes associated with the CHP.[22] The estimates from the modeling studies have been larger than two empirical studies that evaluated the impact of the CHP on outdoor $PM_{2.5}$ in southern, urban regions finding varying reductions (18.6% and 5 $\mu g/m^3$) in outdoor $PM_{2.5}$.[23,24]

1.4 Overview and objectives of this Work

The overall objectives of this work are to evaluate the impacts of the CHP on outcomes expected to be impacted by the policy, apply data-driven methods to develop unique outcomes, and evaluate how the measured and developed outcomes inform the policy and each other. To achieve these objectives, we:

1. Conducted one of the largest, field-based measurement campaigns of a real-world household energy intervention, collecting air pollution and stove use measurements in rural villages near Beijing, China.
2. Deployed machine learning and source apportionment to develop novel outcomes related to stove use (two outcomes) and air pollution (one outcome) to be used in policy analysis models.
3. Implemented recently developed regression and machine learning methods to estimate the impact of the CHP on the measured and developed outcomes.
4. Compared measurement methods and strategies to better understand how to evaluate household energy interventions in the future.

CHAPTER 2: IMPACTS OF A CLEAN HEATING PROGRAM ON STOVE USAGE IN NORTHERN CHINA

2.1 Introduction

Air pollution exposure remains an environmental and public health concern, contributing to respiratory diseases, cardiovascular conditions, and premature mortality worldwide.[2,25,26] Solid fuel combustion for cooking and space heating is a major source of air pollution, especially in rural areas that may lack access to infrastructure for lower-polluting options. In response to the realized impacts of air pollution, various programs and policies have been implemented to accelerate the transition from solid fuel heating towards cleaner alternatives like electric- or gas-powered heating systems. Replacing coal-powered heating stoves with cleaner energy alternatives has the potential to reduce air pollution and improve health outcomes.

Previous studies of solid fuel-to-cleaner energy programs have found no or lower than expected impacts on air pollution and health outcomes.[18,27,28] One potential reason for the unexpected impacts of past household energy interventions and policies on air pollution and health may be due to changes in the use of one solid fuel type when another is targeted for elimination or restriction.[29,30]. Household energy interventions and policies oftentimes target one solid fuel type, but it is common for a household to use multiple types of solid fuel to cook or heat. For example, one study estimated that 89% of homes in Northern China use more than one solid fuel (mostly coal and biomass) to heat their homes in the wintertime.[31–34] If a policy or intervention restricts only one solid fuel type, participants may change their use behavior of the other fuel types, which may mask the impacts of a policy on air pollution. To fully understand the effectiveness of clean heating policies, it is essential to systematically monitor and evaluate supplemental home heating energy use. Without accounting for potential shifts in household fuel use behaviors, assessments of air pollution reductions may be incomplete, highlighting the need for robust methods to track secondary heating stove usage in the context of energy transitions.

The extent to which the use of a supplemental heating stove changes in the context of a policy designed to reduce air pollution from a different heating stove has not been previously evaluated. Past field studies of stove use have been related to energy interventions that target solid fuel burning cookstoves.[35–38] The two most common methods to quantify cookstove usage include hand identifying events in high resolution air pollution or stove temperature data, or hand identifying events in a subset of data and developing an algorithm to classify events in the remainder of the data.[37,39–41] In general, in-home cooking events tend to follow sharp peaks in time series of stove temperature and indoor air pollution since the cookstove is used intermittently. One approach that may prove valuable for identifying stove use events in a large amount of data is machine learning where a model is trained using a subset of visually-identified stove use data and used to predict stove use events in the larger dataset.[36] Unlike rule-based classification algorithms where researchers must define specific thresholds and patterns, machine learning methods develop the classification criteria through data-driven modeling. Leveraging the capacity of machine learning methods to model complex classification systems can potentially improve the accuracy and scalability of stove event classification methods.

Accurately measuring stove use is a resource-intensive process that is generally conducted as part of a larger field study of household energy interventions. These studies often rely on self-reported data from participants about fuel and stove use, which are incorporated into economic and epidemiological models. Previous studies of cookstove interventions that have deployed both survey and stove use monitors to evaluate stove use have found poor alignment between the two methods.[42–44] While these inconsistencies highlight the limitations of self-reported data, most existing studies have focused on cookstoves rather than heating stoves. As a result, little is known about how clean energy policies for heating influence self-reported versus objectively measured stove use. Given the relative scarcity of heating stove interventions, understanding if self-reported and objective measures align in this context is important. If the two measures produce similar results, future studies may be able to prioritize other research efforts over deploying resource-intensive stove use monitors.

To assess and evaluate biomass kang heating stove use, we developed general usage event criteria and trained a XGBoost machine learning model on a subset of manually labeled heating stove use data. Using the predicted biomass kang stove usage, we evaluated the impact of a policy designed to transition villages in Northern China from using coal-powered space heating to electric-powered heat pumps on biomass kang usage. The policy impact was evaluating using an extended two-way fixed effects model for a difference-in-difference study design. We also compared the impact of the policy on objective stove use to self-reported usage to assess their agreement. To our knowledge, this is among the first studies to empirically evaluate the impact of a clean heating policy on non-targeted heating stove usage.

2.2 Methods

2.2.1 Study design and household recruitment

The measurements and data collected for this analysis were part of a larger study to assess the impact of the Clean Heating Program for Northern China (CHP), designed to transition homes from coal stove-based space heating to electric-powered heat pumps, on a variety of environmental, health, and economic outcomes. We enrolled 50 villages from 4 administrative districts (Miyun, Mentougou, Huairou, and Fangshan) in the Beijing region, 20 of which were enrolled in the CHP by the final year of the study (winter 2021-2022). All enrolled villages reported using coal and biomass as their primary heating fuel during the first study year (winter 2018-2019). The villages were located approximately 50 kilometers from the Beijing city center in rural, mountainous regions. Approximately 20 households in each village were approached to be enrolled in our study. Of those 20 homes, ~5 were randomly selected for biomass kang surface temperature measurements. Participant households were required to be occupied during the winter months with no plan to move within the next year. Repeat measurements in subsequent studies waves were collected when possible, and if not available, a new home was selected for biomass kang surface temperature measurement. Temperature measurements were collected in winter 2019-2020 (Wave 2 [W2]), 2020-2021 (Wave 3 [W3]), and 2021-2022 (Wave 4 [W4]).

2.2.2 Biomass kang heating stove use measurements

We measured household biomass kang stove use with iButton (DS1921G-F5; Thermochrom, Maxim Inc.) temperature sensors. Temperature sensors were placed on the surface of the stove and set to take a temperature measurement at the specified time following previous studies.[40] Homes were randomly selected for short-term or long-term measurements. Short-term measurements were conducted over 24-hour periods at high temporal resolution (2-minute intervals) to provide detailed insights into individual heating events. Long-term measurements spanned the duration of the heating season (November-March) at 125-minute intervals and were used to evaluate seasonal trends in stove use. For short-term measurements, we collected 198 in W2 and 134 in W3. For long-term measurements, we collected 137 in W2, 127 in W3, and 183 in W4. We also installed a wall-mounted temperature sensors in each sampled home to understand the baseline indoor ambient air temperature. The indoor temperature was used to distinguish heating stove use from background changes in temperature.

2.2.3 Additional study measurements

We conducted a comprehensive household questionnaire in each study wave to collect a wide range of information related to household demographics, socioeconomic status, stove and fuel use patterns, and behavior related to air pollution exposure. The questionnaire included information about household assets and income that was used to develop a wealth index outlined in Li et al., 2022.[45] We also collected information about home structure including home area and if any new insulation had been installed since the previous study wave. Finally, information was gathered about household heating methods which included estimates of the quantity of each fuel, which heating device was used to heat each room of the home, and the frequency at which heating devices were used to heat rooms.

Village-level population data was transcribed from statistical yearbooks that are published annually by each administrative district. Meteorological information (temperature, relative humidity, wind speed, wind direction, and boundary layer height) was acquired from the European Center for Medium Range Weather Forecasting ERA5 reanalysis.[46] The meteorological data is gridded at 31 km^2, so we used inverse distance weighting to interpolate the meteorological data from the nearest four grid points to the village center. Village-level meteorological values were assigned to all homes within that village.

2.2.4 Stove use event identification

We randomly selected 15% of stoves with measurements to develop general stove use event definitions and to manually label stove use events for both long-term and short-term measurements. The criteria for defining a stove use event were guided by prior research on solid fuel combustion and the primary outcomes of this study, which focused on the air pollution and health impacts of the CHP. Specifically, stove use events were defined in relation to their potential contribution to indoor air pollution, as previous studies of stove use activities in our study region have shown that the addition of solid fuel to stoves is associated with concurrent increases in particulate matter concentrations.[47] Here, we defined stove usage events, generally, as measurable increases in stove temperature across a certain time. The minimum temperature change for an event was 2.5° C between the event start and end with a heating rate of 1° C per hour. Events must have ended at least 5° C above indoor temperature.

Consistent with previous work, temperature data was manually labeled using the TRAINSET online graphical tool which visualizes time series data and allows for each data point to have a chosen label.[48] We labeled stove use event starts and stove use event ends based on imprecise temperature trends. Observations were considered an event start when the preceding observations were decreasing or constant and the proceeding observations increased by ≥ 1° C per hour. Event ends were generally defined as local maxima where the preceding observations increased by ≥ 1° C per hour and the proceeding observations decreased or remained constant. The changes in temperature were generally larger than 1° C per hour, but the criteria were useful when deciding between two sequential observations for a start or end. The criteria were developed based on visual inspection of time series and comparisons to real-time indoor $PM_{2.5}$ concentrations and were developed to minimize misclassification due to ambient changes in temperature. While the criteria were designed to be systematic, some degree of subjectivity was involved due to the complexity of events such as multiple fuel additions during a singular event. To mitigate this, 4 researchers manually labeled events together in 50% of the training data, and separately hand-identified events in the remaining 50% of the training data. The individualresearcher manually labeled data was compared and generally agreed well. In cases of disagreement, discussion occurred and modifications to the criteria were made if needed and applied retroactively. An example of a manually labeled file for a long-term measurement of a biomass kang is shown in Appendix A1.

2.2.5 Long-term biomass kang heating stove use modeling

For this study, we focused on modeling long-term kang use only. Short-term measurements were collected in W2 and W3 only. COVID restrictions in W3 prevented the administration of the household questionnaire. The lack of potential covariates limits the use of the short-term data for policy impact analysis. Instead, the short-term data was used to evaluate if the lower time resolution long-term data captured the same variability in stove usage events and event duration as the higher-resolution short-term measurements. Detailed information about the model training for short-term measurements can be found in Appendix A2.

Half of the long-term manually labeled data was used to train one XGBoost model to predict the probability that each observation was an event start and a second model to predict the probability that an observation was an event end. XGBoost models are built from weak learners (small decision trees) that are trained sequentially from the residuals of the previous weak learner.[49] The outcome variable was the manually labeled event marker (start, end, or neither). The supplied predictor variables were informed by the hand coding process and included the stove temperature as well as the derivative of the lagging and leading data points (± 5 points).

Model performance was evaluated using the F1-score, which incorporates true positives, false negatives, and false positives and can be a better estimation of model performance than sensitivity and specificity for imbalanced data.[50,51] The F1-score incorporates precision and recall compared to AUC (area under the receiver operator curve) which is the measure of predictive model performance that incorporates sensitivity and specificity. Specificity is a measure of non-event predictive performance, which is less useful in data with a large number of non-events relative to events because the number of true non-events will be much larger than the number of non-events identified as events, so sensitivity will remain high regardless of model performance. The F1-score incorporates precision instead of specificity, which is a measure of how well a model predicts true events in relation to misidentifying non-events as events (false positives), which is a more useful measure for imbalanced data. Recall is the same as sensitivity (true positives divided by the sum of true positives and false negatives) which is still useful in imbalanced data because error related to true event identification is most relevant.

We determined a threshold probability for converting the predicted probability that an observation was start/end to a binary variable indicating that an observation was a start/end by selecting the threshold value with the maximum F1-score for each model. The remaining half of the long-term hand-labeled data was used as a test dataset and the F1-score was compared between start and ends models. The model with the larger F1-score was chosen as the "true" model, meaning that every point predicted to be a start/end was considered a true start/end. To balance the number of starts/ends, the observation between two "true" event starts or ends with the highest predicted probability from the lower-performing model of being an event end or start was selected.

Following model prediction, post-model cleaning procedures were applied to refine the event classifications. If two starts or ends were identified in subsequent observations, the later observation was considered the true start or end. When two observations between two starts or ends had the same predicted probability of being an end or a start, the end or start was manually selected (10 cases). We also restricted long-term event length to less than 1050 minutes (17.5

hours) based on the distribution of event length in the hand-identified data. We calculated the number of events per household per day as well as the average duration of stove use per day. Additionally, we explored distributions and descriptive statistics to understand diurnal heating trends in biomass kang usage.

2.2.6 CHP impact modeling

We used a difference-in-difference (DiD) approach to understand the impacts of the CHP on the average number of biomass kang events per household per day and individual event duration.[52] The DiD approach compares differences in the treated group pre- and post-treatment to differences in untreated groups over the same period. Comparing differences over time controls for general baseline difference between treated and untreated groups as well as differences in covariates that change equally over time between the groups. More specifically, the modeling methodology applied in this study was an extended two-way fixed effects (ETWFE) approach.[53] This model is a regression that accounts for staggered treatment timing by including a three-way interaction between the wave that the village was treated, the study wave, and if the village was treated that study wave. The three-way interaction ensured that villages were correctly considered treated or untreated at all time points in the study. Standard errors were clustered at the village level to account for the correlation among measurements taken in the same village. The outcome of this analysis was reported as the cohort specific-treatment effects (treatment effect for villages treated in wave x at time t) and as the average treatment effect for the treated villages (ATT) weighted by the number of treated villages in each treatment cohort.

2.3 Results and Discussion

2.3.1 Household and village characteristics

Summary statistics for households where a long-term kang surface temperature measurement was taken are shown in Table 2.1. Meteorological values were averaged from the average values of the variables for when the long-term kang measurements were taken. Outdoor temperature, wind speed, self-reported house area, and self-reported quantity of wood were similar across all study waves. Relative humidity (RH) was generally higher in W2 (56.5%) than W3 (45.3%) and W4 (46.7%), while boundary layer height was generally lower in W2 (356 m) compared to W3 (441 m) and W4 (460 m). The self-reported percentage of homes using wood for either heating, cooking, or both heating and cooking remained mostly consistent across waves.

Table 2.1 Summary statistics of household and village-level characteristics among homes included in this study with a long-term stove surface temperature measurement on biomass kang heating stoves. [(mean (95% CI) unless otherwise noted]

Study Wave	Wave 2	Wave 3[a]	Wave 4
Number of homes with measurements	138	127	183
temperature (°C)	-3.61 (-3.75 - -3.47)	-5.15 (-5.38 - -4.92)	-5.14 (-5.27 - -5.01)
relative humidity (%)	56.5 (56.2 - 56.8)	45.3 (45 - 45.6)	46.7 (46.5 - 46.9)
wind speed (m s^{-1})	1.73 (1.72 - 1.74)	1.94 (1.92 - 1.96)	1.9 (1.89 - 1.91)
boundary layer height (m)	356 (350 - 362)	441 (435 - 447)	460 (456 - 464)
house area (m^2)	125 (117 - 133)	-	124 (117 - 131)
house area heated (m^2)	77.7 (70.5 - 84.9)	-	85.6 (79.7 - 91.5)
quantity wood (kg)	2260 (2010 - 2510)	-	2190 (1920 - 2460)
use wood for heating only (%)	2.70%	-	2.70%
use wood for cooking only (%)	29.7%		25.9%
use wood for cooking and heating (%)	38.5%	-	43.8%

[a]Due to COVID-19 restrictions, household characteristics and fuel use variables were unable to be collected in wave 3.

2.3.2 XGBoost model performance

The sensitivity, specificity, and F1-score for the XGBoost models to predict the probability than an observation was a biomass kang event start or an event end for the long-term measurements are shown in Table 2.2. A cutoff probability of 0.313 for the start model and 0.373 for the end model was determined by using the cutoff value where the F1-score was the largest. Both the start and end models performed exceptionally well, with F1-scores of 0.914 and 0.897 respectively. In general, the models were more specific (start: 99.52%; end: 99.5%) than sensitive (start: 92.9%; end: 86.8%), indicating they were better at predicting true negative events (no start or end) compared to true positive events (start or end). The models being more specific than sensitive is likely due to the imbalance of event starts/ends relative to non-events in the data. However, F1-scores above 90% indicate that both the start and the end model predicted event starts and ends in the long-term kang data with low error.

Table 2.2 XGBoost model performance for biomass kang stove event models developed using long-term stove surface temperature measurements to identify heating stove use event starts and ends.

	Long-term[a] kangs	
	Start	End
Sensitivity	92.9%	86.8%
Specificity	99.2%	99.5%
F1-score	91.4%	89.7%

[a]Data collected at 2-hr intervals from Nov-March

2.3.3 Kang use description

Distributions of event starts and ends by the time of day for long- and short-term kang measurements are shown in Figure 2.1. The event starts followed a nearly identical, bimodal, distribution across study waves. One event tended to start in the morning around 8 am, with a second event starting around 3 pm. Event ends also followed a bimodal distribution that did not vary by wave, with events tending to end around 12 pm and 7 pm. The average number of events per household per day and event duration by treatment status for long- and short-term kang measurements are shown in Figure 2.2. The average number of events per day was nearly identical for long- (average: 1.01; 95% CI: 0.954-1.01) and short-term (average: 0.920; 95% CI: 0.809-1.03) kang measurements. Events were shorter for short-term measurements (average: 154 minutes; 95% CI: 146-163 minutes) than for long-term measurements (average: 272 minutes; 95% CI: 271-274 minutes). These results taken together indicate that, on average, participants had one event per day that lasted an average of 2.5-4 hours, and most often occurred from 8 am-12 pm or 7 pm-11 pm.

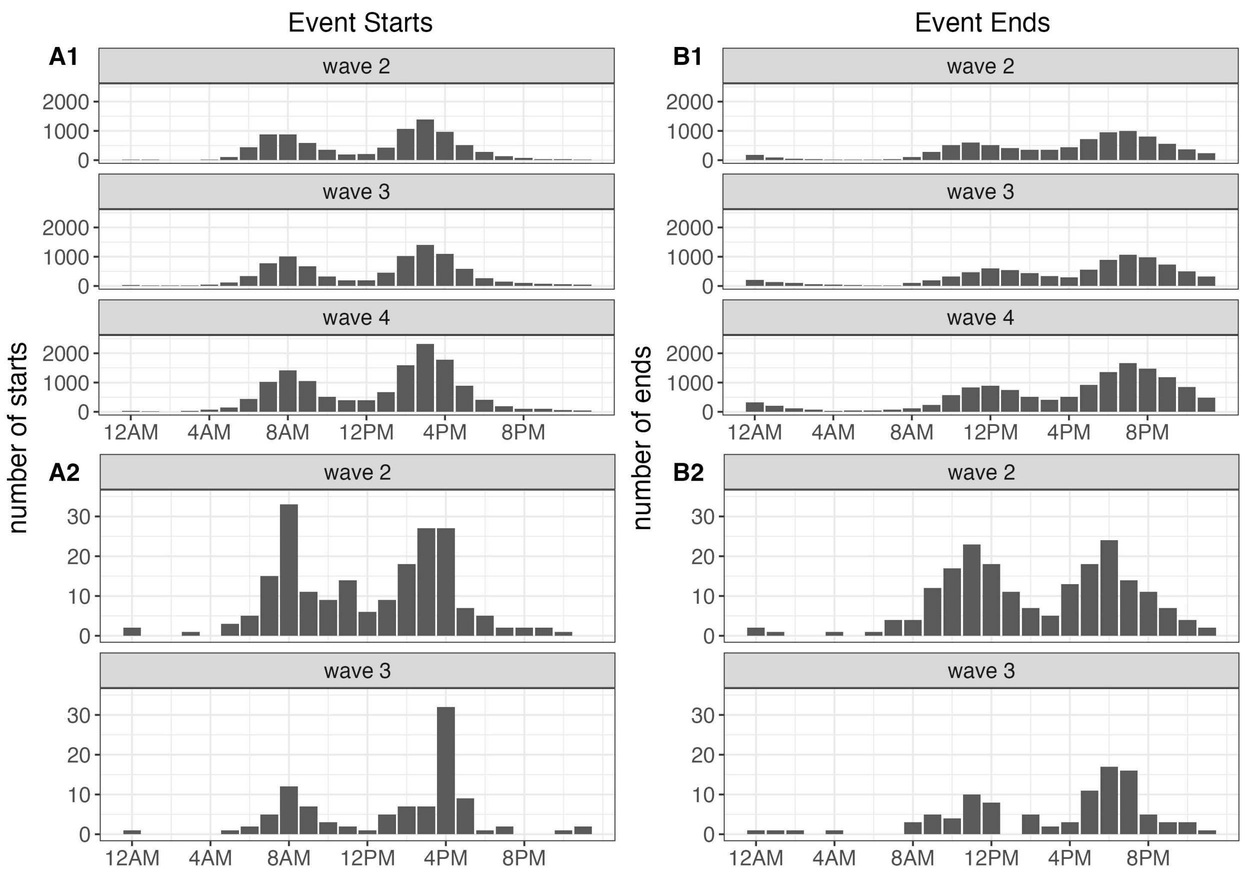


Figure 2.1. Distributions of biomass kang heating stove event starts (Column 1) and ends (Column 2) by time of day and study wave (Waves 2, 3, and 4) for long-term stove surface temperature measurements (A1 and A2) and short-term stove surface temperature measurements (B1 and B2).

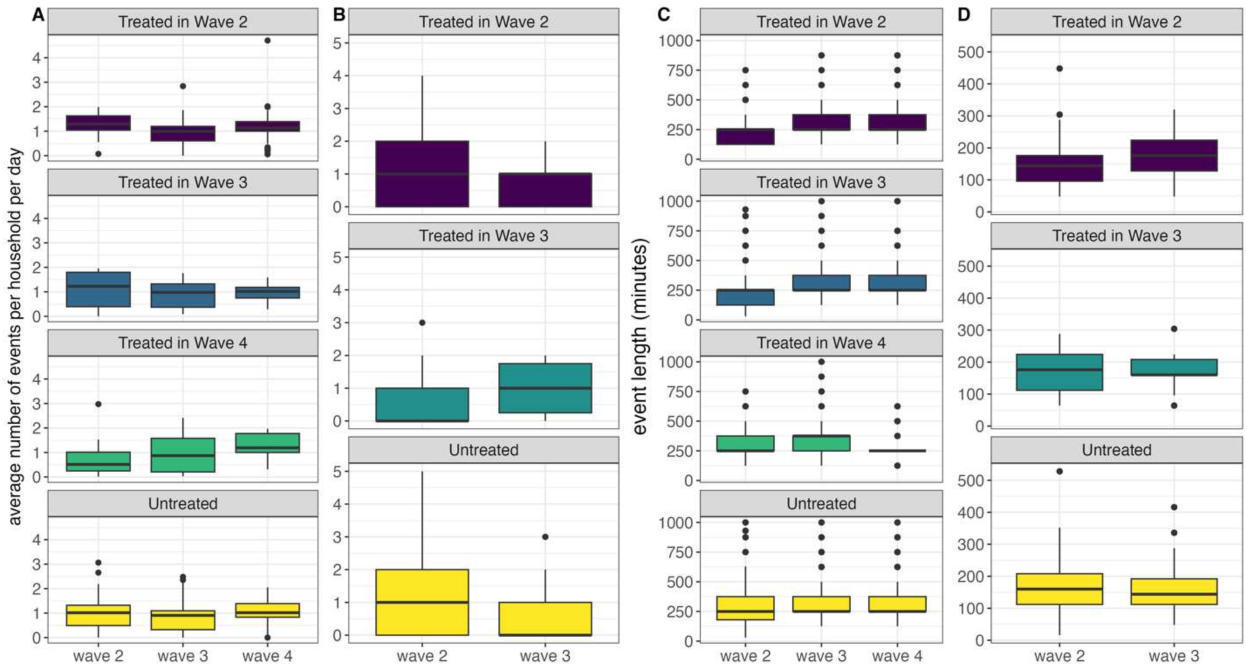


Figure 2.2. Average number of events per household per day for (A) long-term and (B) short-term biomass kang stove surface temperature measurements and distribution of event length for (C) long-term and (D) short-term biomass kang stove surface temperature measurements by Wave and treatment status (four groups: treated in wave 2, treated in wave 3, treated in wave 4, and untreated). In the box plots, the central line represents the median, the box edges denote the interquartile range (IQR; 25th to 75th percentile), and the whiskers extend to the most extreme data points within 1.5 times the IQR. Points beyond this range are plotted individually as outliers.

The one study of stove use that included measurements from biomass kangs (n = 14) found a similar number of events per day (mean = 1.1; SD: 0.6) and event duration (mean: 2.58 hours; SD: 2.58 hours).[36] However, our findings are different than those reported in studies of cookstove events. Ruiz-Mercado et al., found that in rural Guatemala, chimney cookstoves were used an average of 2.56 (95% CI: 2.40-2.74) times per day.[39] In the Tibetan plateau, Clark et al. reported an average of 1.7 (95% CI: 1.6-1.8) events per day for use of a semi-gasifier, traditional wood stove, and both stoves used together.[37] Snider et al., found that participants in Sichuan, China used various cookstoves between 1.7 and 2.4 times per day.[40] The observed differences between heating and cooking stove use align with prior studies indicating that heating stoves tend to have fewer, longer duration events, while cookstoves were used more frequently throughout the day for meal preparation. Divergence of heating stove usage from cookstove usage is an interesting, but expected, finding which shows that heating behaviors are distinct from cooking behaviors.

The distribution of heating stove event starts and ends by time of day and the average number of events per day were nearly identical for long- and short-term measurements. This is in contrast to a cookstove intervention study that found high within-household variability of cookstove use when comparing long- and short-term stove use measurements.[54] However, our results do align with a study that collected surface temperature measurements from cookstoves in Guatemalan homes for alternating 28 day periods over 38 months that found little variability in stove use within households.[39] The diversity of findings across studies suggests that the best measurement duration to capture variability in stove usage is variable by study and should be assessed on a study-by-study basis.

The similarities between long- and short-term measurements indicate that the lower-resolution, long-term measurements captured the same variability in the distribution and number of events as the higher-resolution, short-term measurements. However, the shorter measured event duration for the short-term measurements suggests that the lower time-resolution of the long-term measurements overestimated the average event duration. The comparisons of long- and short-term measurements presented here are useful for future field studies that may involve selecting between long- or short-term measurements. This finding has practical implications for field studies with resource constraints, as it suggests that researchers may choose long-term measurements to reduce the logistical burden of moving sensors every 24-48 hours. Conversely, short-term measurements may be preferable when precise event duration data are required. If a study's primary interest is in estimating the number of heating events, long-term measurements may suffice, while if duration is of greater value, then short-term measurements may be more informative. Our results are specific to a single stove type used in Northern China, so future work is needed to ensure that measurement variability for long- and short-term measurements is consistent across geographic settings, stove technologies, fuel combinations.

Our descriptive findings of kang heating stove usage have some limitations. First, while we trained the short-term models on 15% of the data, the total number of events that the model was trained on was small (n = 64) relative to the long-term models (n = 1628). The small number of events that the short-term models were trained on likely impacted the capacity of the models to accurately predict events and, in part, contributed to the worse model performance for short-term models compared to the long-term models (Appendix A2.2). Future studies that use short-term measurements should hand-label more data for model training. Second, while the average number of events between long- and short-term measurements was similar, the number of short-term events may not necessarily be representative of the number of events at the household level if the home happened to not use the kang during the shorter measurement period. After excluding long-term measurements of kangs in households with no events (n = 13) 18.8% of days across all long-term measurements had no events. If the short-term measurement was taken during a day with no events in a home that had events on other days, then it would not be representative of the long-term measurements at a household level. Future work may opt to take repeated short-term measurements to gain a more accurate representation of events at the household level.

2.3.4 Impact of the CHP on biomass kang usage

The time-varying treatment effects for the impacts of the CHP on the average number of events per household per day and event duration for long-term biomass kang measurements are shown in Figure 2.3. Villages treated in W2 and W3 had small reductions in events per day (~0.1 events per day) that were statistically imprecise, while villages treated in W4 had a small increase in events per day (~0.2 events per day) that was also statistically imprecise. The average event duration did not change due to the CHP, with statistically imprecise increases for villages treated in W2 and W3, and a statistically imprecise decrease for villages treated in W4. No covariates (temperature, relative humidity, wind speed, wind direction, wealth index, or home heated area) included in the models were found to be related to the outcome at an alpha of 0.05. Kangs can be used for both cooking and heating. While we did not collect data regarding the purpose of kang usage, and how often they were used, we did compare stove event start and end times for groups that reported using biomass for only heating and groups that reported using biomass for both heating and cooking and found no relationship between the outcome and biomass usage group.

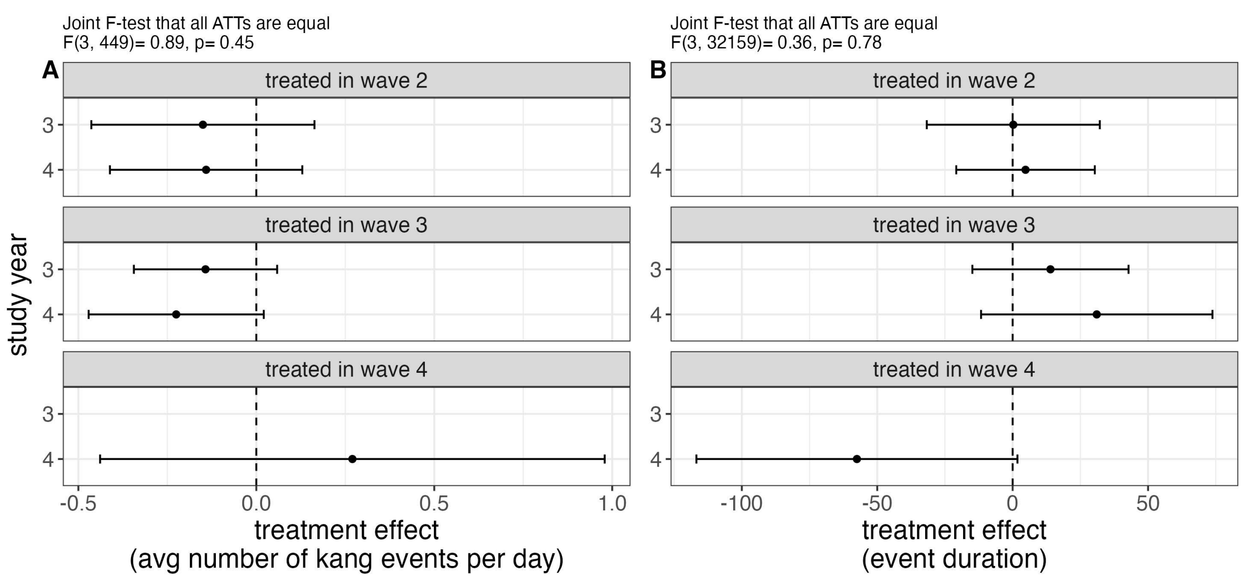


Figure 3. Time-varying treatment effects of the CHP policy on (A) average number of events per household per day and (B) event duration estimated using an extended two-way fixed effects model. No included covariates were associated with treatment status, so the unadjusted results are shown.

Our findings that biomass kang heating stove use did not change due to the CHP are important for future energy transitions. Secondary and stacked cookstove use has been mentioned as a potential reason for the poorly realized health and air pollution benefits of cookstove interventions. This work showed that secondary heating stove use remained the same, which may result in similarly small or no impacts of the CHP on health or air pollution as well as high post-intervention air pollution levels. Future studies should measure the use of all heating devices in the home to understand their contributions to air pollution and provide additional context to the post-intervention air pollution concentrations. Our results are encouraging for future heating energy transitions since secondary heating stove use did not increase in response to the CHP but also shows that more work is needed to eliminate household solid fuel combustion for space heating when multiple heating devices are present.

Our findings that kang stove use and duration did not change due to the CHP has variable alignment with analyses of the impacts of the CHP on self-reported indicators of biomass and kang use (Appendix A3). An analysis of the CHP on total and average heating hours using biomass found that neither outcome changed because of the CHP (Tu et al., in preparation). In contrast, an analysis of the impacts of the CHP on self-reported biomass use (kg of biomass per year) found that biomass use decreased in relation to the policy (average: -487; 95% CI: -805 - -168).[55] Taken together, these results suggest that the frequency and use duration of biomass kang stoves did not change, but the total amount of biomass used at a household level decreased.

Our finding that the self-reported frequency and duration of heating stove events showed similar treatment effects to objective stove-use measures is important for future field studies of heating related energy transitions, particularly to those with budget and logistical constrains. We provide evidence that studies interested in quantifying the impacts of a policy or intervention on stove use can use self-reported stove use data as a reliable measure of stove use. Using self-reported stove use can save the time and cost of deploying stove-use monitors, managing data, and performing the analysis. However, our study did not incorporate the time-series component of the stove use data, which may still offer valuable insight into real-time fluctuations in $PM_{2.5}$ exposure and air pollution dynamics.

Our estimations of treatment effects have several limitations. First, our definition of heating stove usage was rooted in how we expected heating stoves to contribute to air pollution. As a result, we only considered increases in temperature as events, even when the stove was at a constant high temperature and providing heat to the environment. Therefore, we were not able to quantify overall heating usage which may be relevant to other outcomes. Second, we did not analyze a measure of event intensity, such as the difference in temperature between the start and end of an event, which may have been related to total biomass usage and changed because of the CHP. Finally, we did not have information about the purpose of individual kang usage events. Kangs are used for both cooking and heating, so it is possible that kang use for heating reduced while kang use for cooking increased or remained the same.

Notable strengths of our treatment effect estimation include the large sample size of 447 homes, the duration of measurements (November-March), the number of follow-up years, and complementary collection of self-reported data. Studies that have deployed stove use monitors have collected a small number of measurements in few homes, which limits their statistical power and generalizability. Additionally, household clean energy interventions tend to have a pre-post design with no additional follow-up measurements. Our study having one additional follow up year for villages treated in the second and third waves provides some confidence that treatment effect is relatively stable. Finally, our collection and comparison of the treatment effects for stove use measurements to self-reported data provides important insights for future household clean heating interventions surrounding which measurements may be useful to collect.

CHAPTER 3: MULTI-YEAR, FIELD-BASED ASSESSMENT ON THE IMPACTS OF A CLEAN HEATING PROGRAM ON INDOOR, OUTDOOR, AND PERSONAL EXPOSURE TO $PM_{2.5}$ IN BEIJING, CHINA

3.1 Introduction

Air pollution poses a significant environmental health risk, contributing to over 6.5 million premature deaths globally each year.[2,56] Among these deaths, household air pollution from cooking and heating with coal and biomass stoves is responsible for nearly half.[57] Approximately 2.3 billion people worldwide, roughly a third of the global population, and over 100 million rural households in China use solid fuels such as coal or biomass to meet their household energy needs.[57] Consequently, emissions from residential solid fuel combustion substantially contribute to household and outdoor air pollution that disproportionately affects rural populations reliant on solid fuels.[58,59] Transitioning from solid fuel to clean energy is recognized as a strategy to improve air quality and reduce associated health risks.[60–63] Empirical evaluations of household energy transition polices and interventions are important for understanding if implemented technologies and policy levers should be incorporated into future efforts to reduce solid fuel use.

In 2016, the Chinese government implemented one of the largest national energy transition programs in the world to transition households in northern China from using solid fuels to clean energy for space heating.[64–66] Initial studies of the impacts of the Clean Heating Plan for Northern China (CHP) on regional or mostly urban areas have shown promising decreases in outdoor air pollution and modeled personal exposure.[22,24,67,68] Though, modeling studies often assume ideal circumstances, but the real-world implementation of clean energy policies and programs are far more complex. The design and implementation of household energy interventions and policies can have a considerable impact on initial and sustained participation, which may lead to different air pollution impacts than modeled.[18,28] Further, previous studies have been conducted at a city-level spatial resolution using ambient monitoring data. The subsequent lack of information about the impacts of the CHP on indoor or personal air pollution exposure, which is a primary driver for implementation of such policies, motivates the need for evaluations of the CHP that include measures of indoor and personal air pollution exposures. Most previous evaluations of the CHP have been conducted in urban areas, where access to centralized home heating is more common compared to rural areas that rely more on solid fuel for household heating.[69] Given the varying access to household heating options, evaluations of the CHP in rural areas are needed to understand the impacts on air quality in communities that are most impacted by the policy implementation.

This study was conducted to evaluate the impact of the CHP on self-reported solid fuel use and air pollution in a real-world setting using a longitudinal study design. To our knowledge, this is the first study to evaluate the impacts of CHP using field-based air pollution measurements that include outdoor, indoor, and personal exposure. Our findings will provide insights into the actual impacts of the CHP on air quality and personal exposures, air pollution measurement strategies, and the cost effectiveness of the policy. These results will be useful for policymakers and stakeholders seeking to implement or evaluate similar household energy transitions.

3.2 Materials and methods

3.2.1 Study design and participant recruitment

This study was part of the Beijing Household Energy Transition (BHET) study, which aimed to longitudinally evaluate the impacts of the CHP on air pollution and human health. Details about the study are described elsewhere.[70,71] Briefly, we recruited 50 villages from four districts in rural Beijing during the winter of 2018/2019 (Wave-1), with follow-up visits in the subsequent winters: 2019/2020 (Wave-2), 2020/2021 (Wave-3), and 2021/2022 (Wave-4) (Appendix B1). At baseline (Wave-1), villages were eligible but not yet enrolled into the CHP, and households relied on solid fuels (coal and or biomass) for space heating. Between Wave-1 and Wave-2, ten villages were enrolled into the CHP, transitioning from coal to electricity (heat pump) for space heating. Another seven in Wave-3, and three in Wave-4 were enrolled into the CHP, resulting in a total of 20 villages no longer using coal (treated) and 30 still using coal (untreated) by Wave-4 (Appendix B2). Due to the COVID-19 pandemic, a full field campaign was not possible in Wave-3. Therefore, this analysis focused on data collected during Wave-1, Wave-2, and Wave-4. When available, Wave-3 data was included in models as a sensitivity analysis.

In each village, 20 households were recruited to participate based on guidance from village leaders and one eligible adult (> 40 years old) from each household was randomly selected to join our study. In total, we recruited over 1,200 households and 1,400 participants (Appendix B1). All participants provided written informed consent prior to joining the study. This research was reviewed and approved by research ethics boards at Peking University (IRB00001052-18090), Peking Union Medical College Hospital (HS-3184), and McGill University (A08-E53-18B).

3.2.2 Personal exposure measurements

In Wave-1, 10 of the 20 participants per village were randomly selected to wear a personal exposure sampler to collect a 24-h filter sample of $PM_{2.5}$ exposure. In the subsequent waves, we aimed to measure the personal exposure of the same participants as those measured in Wave-1. If a participant was unreachable or declined a personal exposure measurement, we attempted to recruit another adult from the same household. If unsuccessful, we then recruited another participant from another household already enrolled in the BHET study to measure personal exposure. Detailed information about participant recruitment and enrollment is shown in Appendix B1.

Personal exposure samples were collected using personal exposure monitors (PEMs, Apex Pro; Casella, UK) that actively sampled air at a flow rate of 1.8 L/min, and ultrasonic personal air samplers (UPAS, Access Sensor Technologies, Fort Collins, CO, USA) which actively sampled air at 1.0 L/min. Samplers were distributed to participants at random and both housed 37 mm polytetrafluoroethylene (PTFE) filters (VWR, 2.0 µm pore size) and were equipped with a cyclone inlet with a 2.5 µm cutpoint.

3.2.3 Indoor $PM_{2.5}$ measurements

Starting in Wave-2, we randomly selected six households in each village to measure indoor air quality using a commercially available low-cost sensor (PMS7003 Plantower, Zefan, Inc.), which recorded $PM_{2.5}$ concentrations every minute.[71,72] In Wave-4, we aimed to monitor indoor $PM_{2.5}$ in the same households where indoor $PM_{2.5}$ was measured in Wave-2. If a household declined measurements or was unable to be contacted in subsequent waves, we recruited another household enrolled in the BHET study to measure indoor $PM_{2.5}$.

The $PM_{2.5}$ sensor was placed on a table (0.8-1.0m above the floor) in the room where participants reported spending most of their time when awake, such as a living room or bedroom. Indoor $PM_{2.5}$ sensors were deployed between late November and mid-January within study waves 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 after Wave-2 and Wave-4. We also collected co-located $PM_{2.5}$ filter samples (using a PEM or UPAS) in three of the six homes per village, during the first 24-h period of sensor measurements.

3.2.4 Community outdoor $PM_{2.5}$ measurements

To monitor community outdoor $PM_{2.5}$, we set up one to three $PM_{2.5}$ sensors in each village. These sensors were placed at least 500 meters apart and positioned at least 1.5 meters above the ground, away from visible sources of air pollution. Starting in Wave-2, we conducted year-round outdoor $PM_{2.5}$ monitoring with two sensors per village. We co-located a UPAS with outdoor $PM_{2.5}$ sensors in each village to collect filter-derived $PM_{2.5}$ samples. Filter samples were collected and co-located with sensors in rotation in each village, with filters retrieved and replaced every week.

Field blank filters were collected at a rate of ~10% in each Wave for personal, indoor, and outdoor samples, subject to the same field conditions as samples.

3.2.5 Filter analysis

Filters were weighed pre- and post-sampling on a microbalance (Mettler Toledo Inc., XS3DU, USA) after being conditioned at 21–22 °C and 30–34% relative humidity (RH) for at least 24-h and then analyzed for black carbon (BC) using an optical transmissometer (SootScan OT21; Magee Scientific, Berkeley, CA, USA). All filter analyses were conducted at Colorado

State University. The procedures for sample collection, filter preparation, and gravimetric and BC analysis have been described in thorough detail previously.[70,71]

3.2.6 Questionnaire

We conducted a comprehensive household questionnaire to collect information on household demographics, assets, house structure, stove and fuel use patterns, fuel price, and behaviors related to air pollution exposure. The questionnaire can be found on Open Science Framework page (OSF, https://osf.io/8em9f/). Descriptions of how questionnaire responses were used to construct covariates (e.g., smoking exposure) is outlined in Appendix B3.

3.2.7 Outdoor meteorologic data

For indoor and personal exposure measurements, we collected outdoor temperature and dew point data from meteorological stations in Beijing and its neighboring provinces from the National Oceanic and Atmospheric Administration (NOAA) Integrated Surface Data database. For each study home, outdoor temperature and dew point were estimated by inverse-distance weighting the hourly temperature and dew point recorded by government meteorological stations within a 100 km radius of their homes, typically 1 – 4 monitors. Additionally, temperature data were adjusted for altitude using an environmental lapse rate of −6.5°C per 1000 meters for analysis. The twenty-four-hour (the period of personal exposure measurement) mean and seasonal (January 15th to March 15th) mean of outdoor temperature and dewpoint were calculated for each household in three waves. Detailed information on outdoor temperature estimation, as it was conducted for this project, can be found elsewhere.[73]

For outdoor measurements, meteorological data was obtained from the European Center for Medium-range Weather Forecasting ERA5 reanalysis. Village-level daily average meteorological variables were determined by identifying the four surrounding grid points from the reanalysis data and applying inverse distance weighted interpolation from those four points to the village coordinates.

3.2.8 Data processing

Filter samples were excluded from analysis if: 1) the sampling duration of personal and indoor filter samples was less than 80% (19.2 hours) of the target duration (24 hours); 2) outdoor filter sample collection time was less than 12 hours; 3) filter mass was negative or extremely high (e.g., > 2000 μg) indicating a potential error in data collection or data entry; 4) there was missing information on sampling volume of air; 5) filters were damaged including punctures, tears, or holes; or 6) filters were missing during gravimetric analysis and therefore had no mass data. After exclusions, we had 1292, 279, and 695 eligible personal, indoor, and outdoor filter samples, respectively. Descriptive statistics, including arithmetic means and standard deviations (SD), median values, geometric means (GMs) with 95% confidence intervals (95% CI), and minimum and maximum values were determined for indoor, outdoor, and personal $PM_{2.5}$ and BC.

From the low-cost sensor data, we calculated the seasonal mean indoor $PM_{2.5}$ from January 15 to March 15 for each household with indoor measurements, while seasonal average outdoor $PM_{2.5}$ concentrations were calculated from November-February for each village in each wave. A 24-h mean concentration of indoor and outdoor sensor-reported $PM_{2.5}$ was also calculated for the same 24-h period coinciding with the personal exposure measurement period. Due to power shortages, sensor damage, or sensor data loss (Appendix B4), some of our indoor sensors did not continuously monitor indoor $PM_{2.5}$ throughout the entire measurement duration. Sensors that failed to record indoor $PM_{2.5}$ concentrations for more than 80% of the target time period (January 15 to March 15) were excluded from the analysis (52 (17%) in Wave-2 and 69

(23%) in Wave-4). Sensor-reported indoor and outdoor $PM_{2.5}$ concentrations were corrected using the filter-derived $PM_{2.5}$ for indoor and outdoor measurements by wave, respectively (Appendix B5).

Villages were categorized into four groups based on their wave on enrollment into the CHP: (1) *Treated in Wave-2* includes villages that joined the CHP in Wave-2 (2019); (2) *Treated in Wave-3* includes villages that joined in Wave-3 (2020); (3) *Treated in Wave-4* includes villages that joined in Wave-4 (2021); and (4) *Untreated* refers to villages that were not enrolled into the CHP during the study period.

3.2.9 Associations between covariates and air pollution outcomes

A linear mixed-effects model was constructed to identify covariates that contributed to indoor, outdoor, and personal $PM_{2.5}$ and BC levels with participants (for indoor and personal exposure) and villages (for outdoor) as random effects. Fixed effects in models for indoor and personal $PM_{2.5}$ and BC included study wave, outdoor $PM_{2.5}$, outdoor temperature and dewpoint, participant or household smoking status, household heating energy types, wealth index, and participant gender (personal exposure only). Outdoor mixed effects models included the following variables: study wave, outdoor temperature, wind speed, wind direction, and the number of households in each village.

3.2.10 Difference-in-difference analysis

The impact of the CHP on air pollution outcomes was evaluated using a staggered difference-in-difference (DiD) design which is commonly used to more convincingly estimate causal impacts of a policy in observational study settings.[74] The DiD design estimates the impact of the policy by comparing changes in outcomes in treated groups before and after treatment to similar changes in outcomes in untreated groups. Double-differencing is advantageous in this setting because it controls for secular unobserved time-invariant village-level factors as well as secular trends in outcomes affecting both treated and untreated villages. However, this standard DiD model with fixed effects for group and time may be biased when treatment rollout is staggered, as is the case in our study. Therefore, we leveraged recent developments in the DiD literature and implemented an extended two-way fixed effects model to estimate the average treatment effect on treated villages (ATT). The ETWFE model estimated a separate ATT for each treatment group (e.g., the group of villages treated at Wave 2, Wave 3, Wave 4) at time t. We then took the weighted average over the group-time specific estimates to find an overall ATT. Personal exposure and indoor models were adjusted for smoking status, outdoor temperature and dewpoint for personal and indoor air. Outdoor seasonal $PM_{2.5}$ and BC were adjusted for temperature ($PM_{2.5}$ only), wind speed, and wind direction.

3.2.11 Personal exposure and energy consumption

The study calculated the trade-offs between changes in personal $PM_{2.5}$ exposure levels and household energy consumption (coal, wood, and electricity) to further evaluate how the CHP worked to impact air pollution outcomes.[75] Coal and wood consumption were quantified based on the amounts that households reported using during winter, while electricity consumption was calculated from the reported expenses during the same period. Honeycomb coal usage was excluded from the analysis due to its limited adoption among households in this study (<30 households in each wave).

Changes in personal $PM_{2.5}$ exposure levels and energy consumption were analyzed for two pairs of time periods: Wave-1 to Wave-2 and Wave-2 to Wave-4. The analysis included households that participated in both waves for each pair. The average changes in personal $PM_{2.5}$ exposure and energy consumption were then calculated for all villages by treatment status to provide additional context to how the CHP worked. All data analysis in this study was performed using R version 4.2.2.

3.3 Results and discussion

3.3.1 Heating energy use patterns

We observed meaningful transitions from solid fuels (coal and wood) to clean energy (electricity) in villages treated by the policy for all treatment cohorts (Figure 3.1). For example, the proportion of households using electricity exclusively in the villages treated in Wave-2 increased from none in Wave-1 to 24% in Wave-2 and 37% in Wave-4. A similar heating energy source transition trend was observed for households in the villages treated in Wave-3. Notably, households in the three villages treated in Wave-4, compared to the other treated groups, had less willingness to use electricity exclusively. The proportions of households using wood and electricity were dominant (>60%) for all treatment cohorts, which used wood to supplement electric heat pumps. In contrast, most households in the untreated villages kept using coal over the study, while a slower transition from solid fuel to clean energy still can be observed, with the proportions of using electricity exclusively increasing from 3% in Wave-1 to 7% in Wave-2 and 19% in Wave-4 (Figure 3.1).

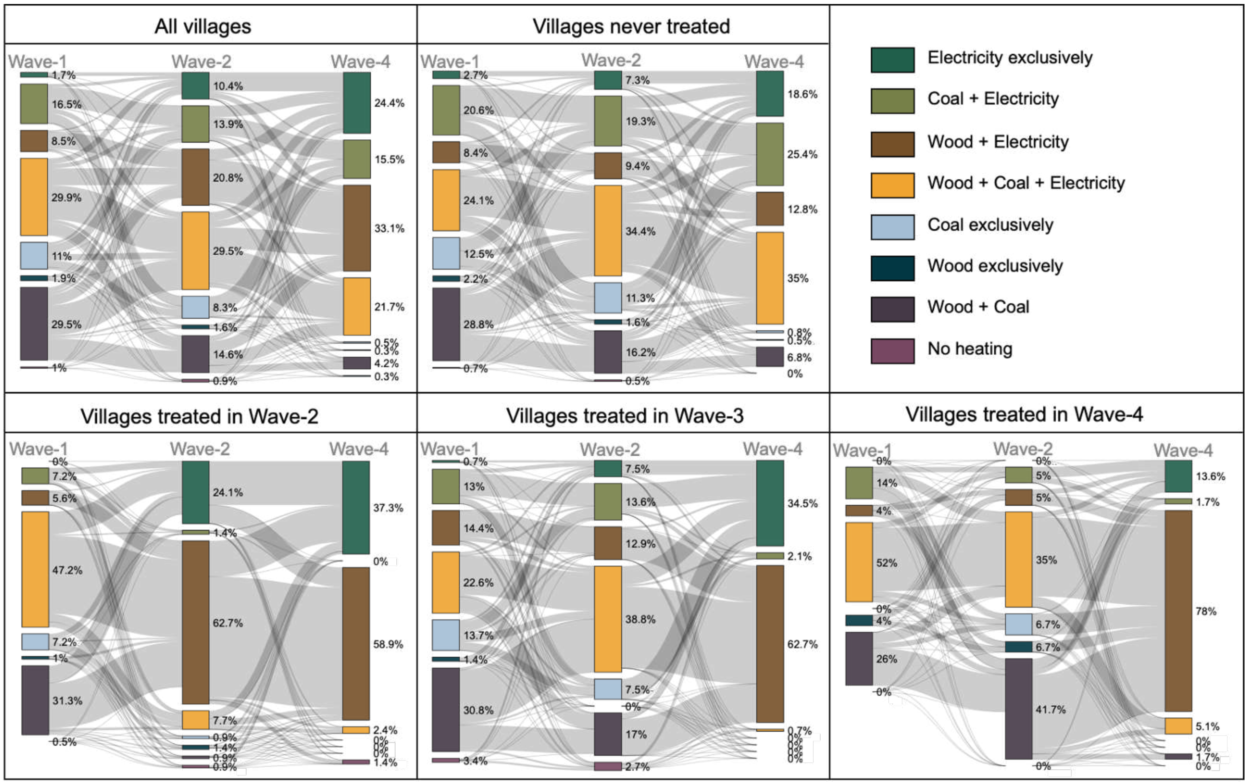


Figure 3.1. Reported household heating fuel types by treatment status across study waves. The color of each bar indicates the type of heating fuel patterns.

We also observed a substantial decline in the self-reported amount of coal (including honeycomb coal) used in the treated villages, from over 2 tons per winter before treatment to almost zero after treatment (Appendix B6). Biomass (i.e., wood logs/twigs), which was usually burned in *kangs* for both cooking and space heating, was not expressly targeted by the CHP policy. We observed declines in self-reported biomass use in villages that were treated in Wave-2 and Wave-3, but there was a small increase in villages treated in Wave-4. In the untreated villages, self-reported use of coal and honeycomb coal also decreased over the study waves, but more slowly than those in the treated villages. Additionally, self-reported use of biomass in untreated villages declined over time. Electricity expenditures increased gradually over time in all villages but rose more sharply in the treated villages compared to the untreated villages. Household spending on electricity in treated villages was approximately three times higher after treatment compared to before treatment. This increase in spending aligns with several studies that have reported similar increased energy costs for rural households resulting from the clean energy transition.[64,76]

The compliance with the CHP observed in our study contrasts with cookstove energy transition projects, where complete compliance has ranged from 14%-64%.[37,77,78] There are several potential reasons for participants' high compliance with the CHP.[55] First, 80% subsidies for night-time electricity rates likely persuaded participants to use the heat pump even if they were concerned with electricity costs. Second, villages were no longer able to purchase subsidized coal, which worked to eliminate the capability to use coal for heating. Finally, all village leaders expressed interest in being enrolled in the policy and encouraged villagers to no longer use coal.

3.3.2 Indoor, outdoor air quality, and personal exposure

Seasonal outdoor geometric mean (GM) $PM_{2.5}$ concentrations decreased consistently across all treated and untreated cohorts from Wave-1 to Wave-4 (Appendix B7). Averaged across groups, the outdoor seasonal GM [95% CI] $PM_{2.5}$ was 49 [44, 54] μg/m^3 in Wave-1, 39 [36, 42] μg/m^3 in Wave-2, and 30 [28, 33] μg/m^3 in Wave-4. Outdoor BC concentrations remained consistently low throughout the study waves (GM < 1.5 μg/m^3), with a slight decreasing trend observed across most groups, except for villages treated in Wave-2.

Consistent decreases in indoor $PM_{2.5}$ concentrations were observed from Wave-2 to Wave-4 (Appendix B7). The GM of sensor-reported 24-h $PM_{2.5}$ in villages treated in Wave-3 and Wave-4 decreased by 60% and 46%, respectively, while sensor-reported seasonal $PM_{2.5}$ decreased by 51% and 49%, respectively. In contrast, smaller reductions were observed in villages treated in Wave-2 and in untreated villages, with decreases of 12% and 30% for 24-h $PM_{2.5}$ and 28% and 30% for seasonal $PM_{2.5}$, respectively. However, indoor BC concentrations increased significantly in villages treated in Wave-3 (doubling) and Wave-4 (tripling), whereas BC levels decreased in villages treated in Wave-2 and those that were never treated.

Personal $PM_{2.5}$ exposures declined markedly in the first year after villages received treatment, with reductions of 31%, 47%, and 37% in villages treated in Wave-2, Wave-3, and Wave-4, respectively (Appendix B7). In untreated villages, the declines in personal $PM_{2.5}$ were smaller, at 27% from Wave-1 to Wave-2 and 15% from Wave-2 to Wave-4. A similar trend was observed for personal BC exposures, though the reductions were less pronounced in villages treated in Wave-3 (only 4%) and Wave-4 (14%).

Mixed-effects models indicated significant reductions in outdoor $PM_{2.5}$ and BC concentrations in Wave-2 and Wave-4 compared to Wave-1 (Appendix B8). Furthermore, high wind speeds and winds originating from the west and northwest were associated with lower air pollution levels. In contrast, higher outdoor temperatures and a greater number of residential households within villages were associated with increased air pollution levels. Increasing use of solid fuel was strongly associated with elevated indoor 24-h $PM_{2.5}$ and BC and personal $PM_{2.5}$ and BC (Appendix B8). This finding suggests that energy transition programs like the CHP should be effective in reducing indoor air pollution and partially improving personal exposures. However, heating fuel use patterns did not show a significant association with indoor seasonal $PM_{2.5}$, which may be related to the reduced use of heating stoves as outdoor temperatures rose toward the end of winter (in March), diminishing the impact of solid fuel combustion to indoor $PM_{2.5}$. Additionally, cigarette smoking was another significant contributor to both indoor and personal $PM_{2.5}$ and BC. Personal $PM_{2.5}$ exposures were positively associated with outdoor $PM_{2.5}$ levels, and females experienced higher personal exposure levels compared to males.

Despite decreasing over time across all untreated and treated cohorts, indoor, outdoor, and personal $PM_{2.5}$ concentrations in our study area remained substantially higher than the health-informed WHO guideline of 5 $\mu g/m^3$. Approximately 60% of participants smoked or lived with a smoker which kept personal air pollution exposure levels high post-intervention. While the quantity of self-reported biomass decreased, a majority of participants in each treatment cohort continued to use their biomass *kang*, which can be a major source of indoor and personal air pollution exposure. Previous studies of household energy interventions have also found that continued use of traditional solid fuel stoves post-intervention as contributors to high post-intervention air pollution levels.[79,80]

Reducing air pollution levels to as low as possible is crucial for improving health outcomes. A meta-analysis of dose-response relationship between long-term $PM_{2.5}$ exposure and various mortality cases showed a relatively constant impact of changes in $PM_{2.5}$ on changes in mortality above 30 $\mu g/m^3$ compared to below the threshold.[81] Reductions in $PM_{2.5}$ above 30 $\mu g/m^3$ likely have a smaller impact on health, which is integral in the justification for air pollution interventions. Post-intervention averages in our study remained greater than 30 $\mu g/m^3$, suggesting that the health benefits of the CHP, mediated by changes in air pollution specifically, are likely to be small. Future air quality interventions should consider a multi-pronged approach to target multiple sources to reduce air pollution to levels where the reduction has noticeable impacts on health.

3.3.3 Difference-in-difference analysis of indoor, outdoor air quality, and personal exposure under the CHP

The results from both the unadjusted and adjusted ETWFE DiD models (Figure 3.2) revealed no significant meaningful impact of the CHP on outdoor winter seasonal $PM_{2.5}$ and BC, as evidenced by the ATT values close to zero. However, the CHP significantly reduced indoor winter seasonal $PM_{2.5}$, with ATTs of -30.9 [-8.7, -53.2] μg/m^3 and -22.2 [-4.2, -40.3] μg/m^3 in the unadjusted and adjusted models, respectively. We also observed a reduction of ~20 μg/m^3 in 24-hr indoor $PM_{2.5}$, though the treatment effect estimation was imprecise (Appendix B9). Conversely, the CHP increased indoor BC (based on filter-derived 24-h measurements), with ATTs of 2.6 [0.4, 4.7] μg/m^3 and 2.9 [0.7, 5.0] μg/m^3 in unadjusted and adjusted models. There was little evidence of any effect of the CHP on personal exposure to $PM_{2.5}$ or BC. This may be due to the reliance on short-term (24-h) measurements, which could capture atypical exposure events rather than representative daily patterns, thereby reducing the precision of policy impact estimations.

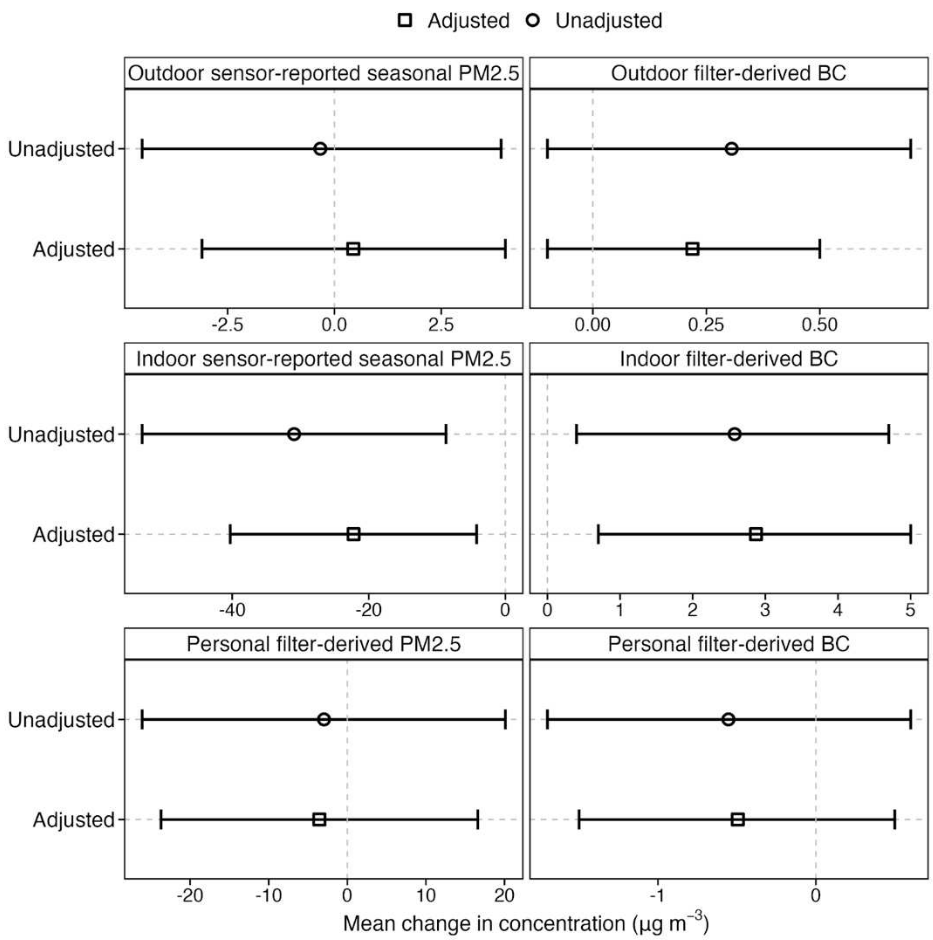


Figure 3.2. The effect of the CHP on outdoor, indoor, and personal exposure to $PM_{2.5}$ and BC using extended two-way fixed effect difference-in-difference analysis (DID). The "*Unadjusted*" y-axis label refers to the ETWFE model without any control variables, while the "*Adjusted*" y-

axis label refers to the ETWFE model incorporating control variables. Dots represent the mean treatment effect, and error bars denote the 95% CIs.

The covariate-adjusted ETWFE estimates were similar to those of the unadjusted models, suggesting a minimal impact of covariate adjustment. We conducted several sensitivity analyses to assess the validity of our findings. First, we examined the number of villages enrolled in the CHP within 20km of our villages to understand if potential air pollution spillover changed differentially between treated and untreated groups. The number of villages enrolled in the CHP remained nearly constant for all treatment cohorts and untreated villages, suggesting that air pollution spillover is, to some degree, accounted for in the difference-in-difference design since it is time-invariant (Appendix B10). We also incorporated a district fixed effect, which was particularly relevant since no villages in the Fangshan district participated in the CHP during the study period, addressing potential differences between the districts that make them poor control units for one another (Appendix B11). Including the district fixed did not impact our treatment estimation, indicating that the untreated villages are good control units regardless of district. We also did not find strong evidence for difference in pre-trends for any outcome, which provides some evidence that the parallel trends assumption is met for later-treated groups (Appendix B12). Our ATT estimations for seasonal indoor and seasonal outdoor $PM_{2.5}$ did not vary with the inclusion of Wave-3 data (Appendix B11).

Our findings on the impacts of the CHP align with the findings of other studies, particularly those examining cooking-related energy transitions on indoor air quality, but differ from findings on personal exposures. For instance, a study in Honduras reported that the use of improved stoves was associated with a 63% reduction in personal $PM_{2.5}$ and a 73% reduction in indoor $PM_{2.5}$ compared to traditional stoves.[82] Baumgartner et al. (2019) observed a complex impact on indoor air quality in a cooking energy intervention study in rural China. After adjusting for differences in outdoor $PM_{2.5}$, they concluded that the intervention reduced winter $PM_{2.5}$ exposures by 46% (95% CI: −70%, −2%). The Household Air Pollution Intervention Network (HAPIN) trial, which conducted one pre- and two post- intervention air pollution and personal exposure measurements in India, Guatemala, and Rwanda, reported significant reductions with LPG interventions: a 92% (95% CI: 90%, 94%) reduction in kitchen $PM_{2.5}$ concentrations and a 74% (95% CI: 70%, 79%) reduction in personal $PM_{2.5}$ exposures.[62]

Two previous studies have estimated its impact on outdoor $PM_{2.5}$ in urban areas. One study reported an 18.6% reduction in $PM_{2.5}$,[65] while the other found a reduction of 5 $\mu g/m^3$,[83] indicating varying impacts of the policy. Despite improvements in air quality observed in our study—both attributable and not attributable to the CHP—pollution concentrations remained significantly higher than the WHO recommended exposure guideline of 5 $\mu g/m^3$. These findings highlight the need for further interventions to meet global air quality standards.

Our study underscores the importance of measurement location and duration for understanding the comprehensive impacts of a household energy transition on air quality. We found different impacts of the policy for different measurement location, indicating that measurements from one location (indoor, outdoor, personal) cannot be used to estimate the policy impact at a different measurement location. The comparison of the 24-h and seasonal average indoor measurements show that, for indoor specifically, longer measurement durations can help improve statistical precision of treatment effect estimation. Since our study, advancements in personal exposure measurement instrumentation have been made that could be used in future studies to improve understanding of personal exposure. For example, recent developments in sensor technology have included real-time $PM_{2.5}$ measurements which, coupled with high-resolution GPS, could be used to further understand personal exposure and potentially control for location-based differences between participants and study waves that are driving variability in $PM_{2.5}$.[84,85] Additionally, future studies could circumvent the 24-h limitations by collecting multiple 24-h measurements per participant per wave to gain a better understanding of the seasonal average personal exposure, or collect more personal exposure measurements to increase statistical power. The authors do acknowledge the cost, time, and logistical challenges of improving personal exposure measurements.

3.3.4 Personal $PM_{2.5}$ exposure and energy consumption trade-offs

The personal $PM_{2.5}$ exposure and energy consumption trade-offs analysis revealed that in the year villages joined the CHP (treated year), reductions in coal consumption were nearly linearly associated with decreases in personal $PM_{2.5}$ exposure levels, while electricity expenditures increased by over 2,000 RMB on average (Figure 3.3). On average, for every 10 $\mu g/m^3$ reduction in personal $PM_{2.5}$ in these villages, households used 2.7 fewer tons of coal and spent an additional 4888 RMB on electricity. In contrast, changes in wood consumption were untraceable due to the policy's exclusion of wood usage for *kang* heating. In villages not enrolled in the CHP, coal consumption also declined, but the reduction in personal exposure levels was achieved more efficiently: every 10 $\mu g/m^3$ decrease in personal $PM_{2.5}$ exposure was associated with 1.3 fewer tons of coal and only 663 RMB in additional electricity cost. Furthermore, changes in wood consumption were proportional to changes in personal exposure levels, emphasizing the importance of wood burning as a contributor to exposure. In treated-treated villages, energy consumption showed minimal change, as households had already adapted to the policy, and personal exposure levels remained relatively stable.

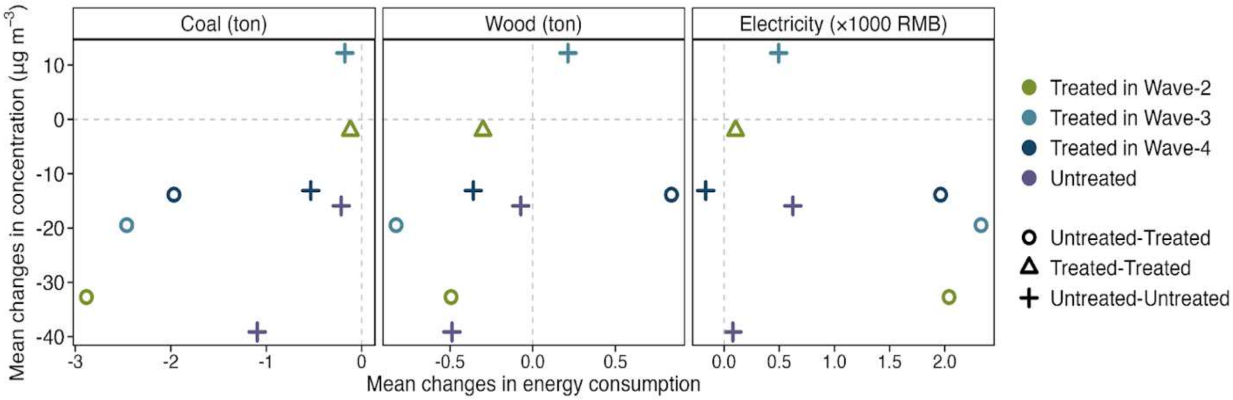


Figure 3.3 Changes in personal exposure to $PM_{2.5}$ relative to changes in heating fuel consumption by type. The shapes of the points represent changes in village treatment status between Wave-1 to Wave-2 or Wave-2 to Wave-4: circles for villages transitioning from untreated to treated, hollow triangles for villages consistently treated, and plus signs for villages consistently untreated. Point colors correspond to the year villages received treatment.

Households enrolled in the CHP showed significantly reduced coal use and increased electricity expenditures as indicated by the high compliance of households in these villages. Similar trends were also observed in untreated villages, albeit at a slower energy transition pace. Notably, untreated households demonstrated higher cost-effectiveness, with smaller reductions in coal use, less increase in electricity expenditure, but comparable personal $PM_{2.5}$ reductions. This suggests that fully restricting coal use for residential heating and relying solely on electricity may not yield optimal air pollution benefits. Future studies should investigate an optimal mix of electricity, coal, and wood for heating to achieve maximum air pollution reductions with minimal additional expenses.

Assessing the impact of the CHP on air quality remains challenging; however, our study stands out as one of the most comprehensive field-based, multidisciplinary investigations to date, incorporating a wide range of air quality measurements across a regional scale and multi-year field campaigns. Several strengths bolster our findings: first, the quasi-experimental difference-in-difference design allowed us to control for time-invariant and secular trends, which strengthens the causal impacts of the policy on outcomes; second, a sufficiently large sample size and multi-year measurements enabled robust pre- and post-intervention comparisons; third, diverse air pollution data, including indoor, outdoor, and personal exposure, were collected, incorporating time-integrated short-term and time-resolved long-term observations, which can help improve statistical precision of treatment effect estimation; finally, detailed covariates, such as household fuel use and demographic information, were obtained through comprehensive questionnaires. We note that controlling for secular trends was important in our case, as we would have found large, significant decreased in personal $PM_{2.5}$ exposure had we not accounted for a time fixed-effect (Appendix B11).

While our study offers valuable evidence for policymakers and other stakeholders aiming to implement or evaluate similar household energy interventions, several limitations must be acknowledged. First, the absence of indoor $PM_{2.5}$ measurements in Wave-1 limited our ability to track changes in indoor $PM_{2.5}$ levels from baseline for the cohort with the most treated villages (villages treated in Wave-2). We were also unable to evaluate pre-intervention trends for villages treated in Wave-2, so our treatment estimation may be biased by differences in pre-trends. Third, the 24-h duration for personal exposure measurements may not fully capture all relevant exposure events, potentially affecting the policy impact estimations. Addressing these gaps in future research could strengthen the understanding and effectiveness of household energy interventions like the CHP. Last, an analysis of changes in source contributions, a topic explored in the broader BHET study but not presented here, could yield additional insights into the coal-specific impacts of the CHP.

A promising transition from coal to electricity was observed in our study area with the implementation of the CHP, evidenced by a significant decrease in coal usage and a notable increase in electricity expenditure. We observed a large decrease in indoor seasonal $PM_{2.5}$.

However, additional effort is needed to improve personal exposure as the indoor air pollution benefits did not translate to observed benefits to personal $PM_{2.5}$ exposure, as the indoor air pollution benefits of the CHP did not lead to subsequent improvements in personal exposure. We demonstrated how distinct air pollution measurements should be taken to understand the comprehensive and heterogenous impacts of the policy on outdoor, indoor, and personal air pollution exposures. Our study highlights the importance of continuous learning and refinement of similar policies to extend their benefits across other regions in China and globally. By sharing lessons learned and offering recommendations for best practices, we can accelerate progress toward improving air quality and achieving Sustainable Development Goals (SDGs) at a broader scale. Collaboration and knowledge exchange will be crucial in driving impactful change and fostering sustainable development worldwide.

CHAPTER 4: SOURCE CONTRIBUTIONS AS AN OUTCOME FOR EVALUATING THE IMPACT OF HOUSEHOLD ENERGY TRANSITIONS: AN EXAMPLE FROM A CLEAN HEATING PROGRAM IN BEIJING, CHINA

4.1 Introduction

Exposure to air pollution from household solid fuel combustion generated by cooking and heating is a leading environmental risk factor for premature death and a wide variety of diseases.[5,86,87] Cleaner alternatives to solid-fuel based cooking or heating devices, including both standardized stoves and those customized to specific communities or use cases, have been shown in laboratory studies to reduce air pollutant emissions.[88–90] However, previous studies of real-world household energy transitions for cooking or heating have found limited or no evidence that the intervention had the intended outcomes on air quality or health.[91,92] Studies of household energy transitions most commonly measure total $PM_{2.5}$, which is a complex mixture of air pollution that originates from multiple sources including combustion, dust, vehicle emissions, tire wear, and secondary aerosols. The chemical composition, chemical properties (e.g., vapor pressure, partitioning, oxidation state), toxicity, and interaction with the physical environment (e.g., meteorology) all vary by the source of $PM_{2.5}$.[93–95] This variability, combined with the fact that most governmental policies and research interventions target one specific source of pollution, makes it challenging to isolate the effects of a policy on an outcomes like air pollution that results from multiple contributing sources.

Some studies of household energy interventions have used chemical components of $PM_{2.5}$ as tracers for sources identified from lab-measured source profiles or factor analysis.[95–98] However, this approach may not work in areas where tracers are emitted from multiple sources.

For example, levoglucosan, a common tracer for biomass combustion, has been shown to also be emitted in significant quantities from coal combustion, reducing its reliability as a tracer of biomass combustion in areas where coal is also used.[99] A source apportionment approach, which apportions the mass of chemical components to modeled sources, has been useful for identifying sources in settings where tracers may originate from various sources.[100,101] Source apportionment methods that identify the source contribution targeted by a source-specific policy intervention could provide a more accurate basis on which to evaluate the impacts of a policy than total $PM_{2.5}$.

New machine learning methods have been developed to estimate the treatment effect of policies in observational settings, and have been shown to accurately estimate the treatment effect using simulated data.[102,103] Modeling non-linear outcome-covariate relationships and covariate interactions is inherent to the structure of machine learning models, which can be advantageous in air quality settings where the relationships are unclear or difficult to specify. While these methods have shown promise, there are limited evaluations of their treatment effect estimation compared to more conventionally accepted estimation methods in a diverse range of real-world settings including air quality interventions, varying levels of treatment assignment (e.g., individual or group level), staggered treatment rollout, and longitudinal study designs. Analyses of how machine learning methods perform in various settings is needed to better understand the strengths, limitations, and usefulness of the approaches compared to conventional estimation methods.

We evaluated the impacts of a Clean Heating Program (CHP) in Northern China on the coal containing source contribution in outdoor and personal $PM_{2.5}$ exposure and compared them to the impacts of the same policy on total $PM_{2.5}$. We also compared the treatment effect estimation of a conventional treatment effect estimation method (extended two-way fixed effect model) to two machine learning methods. To our knowledge this is the first study to use source apportionment methods to evaluate the impacts of a clean heating policy on policy-specific source outcome. This is also the first study to use compare machine learning methods to traditional treatment effect estimation methods on air pollution outcomes in the context of a household energy intervention.

4.2 Materials and methods

4.2.1 Study area and participants

The CHP is a policy set forth by the Chinese government with an aggressive goal to transition 70% of homes to electric or natural gas-powered heating from coal-based space heating. The government provided villages with subsidies to buy and install heat pumps, as well as to cover electricity costs. The CHP started in 2016 and was implemented in yearly waves at the village level.

In the winter of 2018-2019 (Wave-1), we enrolled 50 villages to participate in a multi-year longitudinal study. The villages were located in 4 administrative districts (Fangshan, Hiaurou, Mentougou, and Miyun) approximately 1.5-2 hours from the Beijing city center. The villages were in a generally rural, mountainous region with few local sources of air pollution besides household solid fuel combustion for heating and cooking. In Wave-1, all villages reported using coal as their predominant heating fuel. By the final study year (winter 2020-2021; Wave-4), 20 villages entered the CHP.

We randomly selected 1 participant in ~10 homes per village that were occupied in the wintertime to participate. Participants were required to be living in the village, over the age of 40 and not planning to move within the next year. When possible, the same participants were visited each wave. If a participant was not available in a later study wave, a new member of that household was recruited. If a new member of the household could not be recruited, another household in the village was recruited. Due to COVID-19 restrictions, we were unable to collect key measurements in winter 2019-2020 (Wave-3), so the subsequent analyses do not include data from Wave-3.

4.2.2 Air pollution measurements

$PM_{2.5}$ time-integrated filter measurements were collected outdoors in each village and for ~10 participants per village (November-February) on Zeflour or Teflon filter media (PTFE 37 mm with 2 μm pore size, Pall Labs).

Outdoor measurements were collected over the course of 7-14 days for the duration of the field campaign using an Ultrasonic Personal Aerosol Sampler (UPAS, Access Sensor Technologies, CO, USA) operated at a flow rate of 1 L min^{-1}.[104] The UPAS was placed inside a custom-built enclosure designed to prevent resampling air and to protect from environmental damage. We collected 121 samples in Wave-1, 349 in Wave-2, and 247 in Wave-4 and ~10% field blanks per wave. Due to limitations in data collection in the first study wave, only 42 villages were sampled in Wave-1 while 50 were sampled in Wave-2 and Wave-4.

We also collected co-located measurements on quartz fiber filters for a subset of Teflon filters outdoors and indoors to determine elemental and organic carbon. The outdoor quartz filters were collected with a UPAS over 7-14 days operating at a flow rate of 1.8 L min^{-1} and the indoor quartz filters were collected over a 24-hr period using Personal Exposure Monitors (PEMS, Harvard) operated at 1 L min^{-1}. The quartz filters were baked at 550 °C for a minimum of 8 hours prior to sampling to remove any impurities. Outdoors, we collected 23 co-located samples on quartz filters in Wave-2, and 9 in Wave-4 with 3 field blanks in both waves. Indoors we collected 57 co-located samples on quartz filters in Wave-2 (14 blanks) and 102 in Wave-4 (18 blanks).

Personal exposure measurements were collected over a 24-hr period once per field campaign for ~10 participants per village using a UPAS or PEM. The UPAS collected air at a flow rate of 1 L min^{-1} while the PEMs sampled air at 1.8 L min^{-1}. Sampler flow rates were calibrated the evening prior to deployment and were measured immediately after the sampling period. Participants were instructed to always wear the small waist pack (for the PEM) or the lanyard (for the UPAS) except when sleeping, showering, or sitting when they could place the sampler close by. We collected 422 samples in Wave-1, 367 in Wave-2, and 369 in Wave-4 and ~10% field blanks per wave.

4.2.3 Air pollution sample analysis

Before and after sampling, filters were weighed in triplicate using a microbalance (Mettler Toledo Inc. XS3DU, USA) after being conditioned in a temperature (21-22 °C) and relative humidity-controlled environment (30-34% RH) for 24-hrs. They were then weighed in triplicate (1 μg resolution), and the average of the three measurements was used as the mass.[105]

Mass concentrations of 12 individual elements were measured by energy-dispersive X-ray fluorescence (Thermo Scientific Quant'X EDXRF).[96] The filters were then sectioned in half and extracted in water to quantify eight water-soluble ions via ion chromatography.[106] Quartz filters were analyzed via thermal-optical EC/OC analysis (NIOSH method 5040; Sunset Laboratories EC/OC analyzer).[107] Teflon filters were scanned for color-scale parameter using a colorimeter (i1Pro Colorimeter, X-Rite Inc).[108] A random forest model was trained to predict the thermal-optical EC/OC concentrations as a function of the color-scale parameters for co-located

Teflon and quartz measurements (~200 measurements). The random forest model was then used to predict EC/OC concentrations on the remaining Teflon filters with no co-located quartz filter.

The water-soluble (ws) and water-insoluble (wi) fractions of species were found by subtracting the water-soluble component measured by IC from the total elemental concentration measured by XRF in cases where species were measured by both IC and XRF. Non-sulfate sulfur (ns-S) was determined by subtracting the amount of sulfur found in the sulfate measured by IC from the total elemental sulfur concentration measured by XRF.

4.2.4 Source apportionment

The sources and their contributions to outdoor and personal $PM_{2.5}$ exposure were found using the EPA Positive Matrix Factorization (PMF) 5.0 model. PMF is a factor analysis model that uses chemical speciation data to determine the source profiles (the concentration of each chemical species per unit of each source) and their contributions (the concentration of each source in each sample).[109] PMF identifies the optimal sources and their contributions by minimizing the squared error (Q) from a general chemical mass balance formula that is scaled by the uncertainty associated with the chemical speciation data. The optimal number of sources was assessed using the PMF diagnostic tools. Reductions in the change of in squared model error (Q) divided by expected squared model error ($Q_{expected}$) moving from lower to higher solutions indicate that adding additional sources is overfitting the model. $Q_{expected}$ is calculated based on the number of sources in the solution and the number of species with a high signal-to-noise ratio. Displacement swaps (DISP swaps) assess rotational error by making small changes in the source profiles, calculating the remaining source profiles and contributions that result in a small change in Q, and determining if the new source profiles differ from the original. If DISP swaps occur, the solution is unstable. Bootstrap mapping runs models on resampled data, compares the new source profiles and contributions to the original model, and determines how correlated the new model is with the original. Sources that have low bootstrap mapping, meaning they are not consistently present in the bootstrapped models, indicates that the given n-factor solution is unstable.[110]

Outdoor and personal $PM_{2.5}$ exposure measurements were included in the same PMF model and the optimal number of sources was determined using the PMF model provided solution selection tools described in the previous paragraph. The chemical composition data used as the input for the PMF model were dispersion-normalized using wind speed and boundary layer height.[111] As a result of the dispersion normalization process, wind speed and boundary layer height were not adjusted for in the subsequent policy analysis models since they were accounted for in the outcome.

4.2.5 Other study measures

For each participant, field staff administered a questionnaire during each study wave to collect information about demographics, socioeconomic status, and behaviors that may contribute to indoor air pollution or air pollution exposure. Hourly meteorological data (boundary layer height, temperature, relative humidity, wind speed and direction) were interpolated to the village level via inverse-distance weighting using data from the European Center for Medium Range Weather Forecasting ERA5 reanalysis. Village-level population and number of households were transcribed from district-level statistical yearbooks. A list of all villages in the Beijing area as well as the year they entered the CHP was obtained from the College of Urban and Environmental Sciences at Peking University and matched to coordinates using data from the China National Bureau of Statistics for Wave-1 and Wave-2. For each of our study villages, we calculated the total number of villages, and the proportion of villages enrolled in the CHP within 20km for a sensitivity analysis to understand the potential impacts of spillover on outdoor air pollution.

4.2.6 Statistical analyses

We estimated the average treatment effect for villages that were enrolled in the CHP (average treatment effect for the treated; ATT) using an extended-two-way fixed effects (ETWFE) model for the individual-filter level source contributions to the coal-containing source with standard errors clustered at the village level.[53] This method compares changes in treated villages to changes in untreated villages over time. Comparing changes over time accounts for fixed differences in treated and untreated villages and covariates that change uniformly over time in both groups. We also present the time-varying treatment effect for each cohort group at each wave treated (e.g., villages treated in Wave-2 in Wave-4) by specifying a three-way interaction between wave treated (treatment cohort), study wave, and whether the village was treated in the given study wave. Covariates considered for the ETWFE and machine learning models were outdoor relative humidity, temperature, boundary layer height, wind speed, the number of homes in a village, the village population, gender, smoker, nonsmoker living with smoker ("indoor smoking"), and self-reported wood use.

The first machine learning method we applied is dynamic causal forests (DCF).[102] Similar to random forest models, causal forests are an ensemble method that builds a series of causal regression trees and averages the predictions across trees to make the final prediction of the treatment effect. At each split, the conditional average treatment effect for the treated (CATT) is calculated for a subset of randomly selected covariates with the variable to split on and the splitting rule determined by minimizing a function to estimate the expected mean squared error (EMSE) which consists of two parameters. The purpose of the EMSE function is to balance the bias-variance tradeoff associated with model complexity (depth of trees in this case). The first parameter accounts for the difference in CATTs between splits. By maximizing this parameter, the model favors splits that have different CATTs, reducing variance. The second term minimizes the within-split variance (by subtracting this value from the previous parameter) to prevent overfitting and reduce bias.

The second machine learning approach we applied to evaluate the impacts of the CHP is Bayesian causal forests (BCF) outlined in Hahn et al., 2020.[103] This method is an extension of Bayesian additive regression trees (BART), which function like a XGBoost model where small regression trees (weak learners) are trained on the residuals of the previously built tree. The final prediction is the sum of the predictions from the weak learners with some noise added. BCF incorporates two separate models. The first model assesses the influence of covariates on the outcome by incorporating the propensity score as a covariate to account for confounding. The second model evaluates the impact of the treatment effect conditional on the covariates. By modeling these two parameters separately, BCF allows regularization of both the covariates and the treatment effects. The regularization in BCF is achieved through the selection of prior distributions for the splitting rules that heavily penalize trees with many splits.

To implement a staggered difference-in-difference design with DCF and BCF methods, we used the difference in the village-level average outcome between the treated year and the most recent year untreated in separate models for each treatment cohort.[112,113] We included the covariate levels for the most recent untreated year, as well as the difference between the treated year and the most recent untreated year. The overall study ATT was estimated by summing the mean and 95% confidence (for DCF) and credible (for BCF) intervals from the cohort-level models weighted by the proportion of treated villages in each model.

4.3 Results and discussion

4.3.1 Participant characteristics and policy uptake

Table 4.1 shows the arithmetic mean and standard deviation for select demographic and environmental measures for never treated and ever treated groups as well as the difference in means at baseline (Wave-1). The never treated and ever treated villages were similar for age, percentage male, percentage of smokers, percentage of indoor only smoking, and all the meteorological parameters. At baseline, villages that were treated had a higher percentage of the population that used wood (79.0% vs 61.5%) compared to never treated groups. Villages that were treated also had a lower average population (188 vs 282) and lower number of households (432 vs 619) compared to never treated villages.

The policy uptake for our study was exceptionally high, with 0% of participant households using coal exclusively and 7.40% using coal in addition to another fuel source (wood and or electricity) in the year immediately following policy implementation for treated villages. Coal use in the never treated villages remained relatively stable, with 78.4% (SD: 9.32%) of never treated villages using coal throughout the study duration. The high policy uptake in our study is uncommon, with previous cookstove interventions studies indicating low compliance.[114–116] Potential reasons for the high uptake in our study include the relative ease of using heat pumps compared to having to continuously add coal throughout the day, subsidized electricity, and general support for the policy indicated from village-committee surveys.[55]

Wood, exclusively or in addition to another fuel, remained a popular fuel source for all homes enrolled in our study. Across all study waves, 80.5% (SD: 5.75%) of homes reported using wood, the majority of which used wood in addition electricity and or coal. Wood and electricity was the most common fuel combination post-treatment, with 62.7% of homes treated in Wave-2 and Wave-3 and 78% of homes treated in Wave-4 having self-reported both wood and electricity use. While coal use effectively ceased in villages post-treatment, our self-reported measures indicate that a solid fuel source was still being used for household space heating. The continued use of a solid fuel source likely contributed to high post-treatment air pollution levels in our study.

Table 4.1. Descriptive statistics for demographic and environmental measures at baseline by treatment status.

	Never treated (outdoor n = 80; personal n = 265)	Ever treated (outdoor n = 41; personal n = 157)	Difference in means
	Mean (SD)	Mean (SD)	Difference (SE)
Demographics[a]:			
Age	60 (9.43)	60.7 (8.5)	0.676 (0.916)
Male (%)	42.3 (49.6)	38.2 (49)	-4.05 (4.98)
Smoker (%)	26.4 (44.2)	26.8 (44.4)	0.336 (4.46)
Indoor smoking[c] (%)	32.8 (47)	32.5 (47)	-0.346 (4.74)
Used wood (%)	61.5 (48.7)	79 (40.9)	17.5 (4.63)
Use any coal pre-intervention (%)	78.4 (9.32)	86.1 (8.25)	-7.70 (7.12)
Use any coal post-intervention (%)		7.40 (4.16)	
Environmental measures[b]:			
Temperature (°C)	-4.42 (2.73)	-4.76 (3.33)	-0.343 (0.566)
Relative humidity (%)	35 (8.29)	34.4 (8.8)	-0.643 (1.63)
Boundary layer height (m)	345 (98.6)	350 (104)	4.56 (19.3)
Wind speed (m s^{-1})	1.86 (0.348)	1.87 (0.336)	0.0059 (0.0661)
Number of households	282 (215)	188 (114)	-94.2 (35.9)
Population	619 (509)	432 (301)	-187 (86.4)

[a]Taken from personal exposure measurements
[b]Outdoor village-level averages
[c]Participant lived with a smoker but was not a smoker themselves

4.3.2 Source apportionment

Arithmetic mean and 95% confidence intervals for chemical constituents included in the source apportionment model are shown in Appendix C1. Positive Matrix Factorization (PMF) model diagnostics are summarized in Appendix C2. We observed the largest reduction in $Q/Q_{expected}$ moving from three to four factors (0.87) while the changes in $Q/Q_{expected}$ were similar moving from four to five (0.53) and five to six (0.54) factors. The minimal change of $Q/Q_{expected}$ moving from four to five factors suggested that the four-factor solution was the most optimal. There was little rotational error (DISP swaps) in any of the solutions (Appendix C2). The new sources identified in the five- and six-factor solutions (Pb-dominated and chloride-dominated factors) had poor bootstrap mapping, which suggests that the additional factors after four did not result in stable solutions. Based on the diagnostic criteria, the four-factor solution was selected as the final model.

The source profiles and contributions for the four-factor PMF solution are shown in Figure 4.1 and Figure 4.2. Differences in the percentage of source contributions to the sum of the source contributions (SSC) between outdoor and personal exposure were evaluated by a two-sample t-test. The first source was identified as dust from the high percentages of common crustal elements like Si, wi-Ca, and wi-Mg. Dust contributed a higher percentage of the SSC for outdoor (mean: 14.5%; 95% CI: [11.5-17.5%]) compared to personal exposure (11.8% [9.26-14.4%]). Dust is the most common factor identified in source apportionment models globally and tends to be the predominant source of non-ionic metals in China. In our study, dust contributed a similar percentage of the SSC as dust apportioned from personal PM exposure samples collected in urban Beijing and suburban Huairou (16.5% ± 15.9%).[117] Outdoor dust in our study contributed more to the SCC than dust apportioned from outdoor PM samples collected at a regional background site in northern China (7.24%), which is likely due to increased activities in the villages that would suspend or resuspend dust relative to an isolated field location.[118]

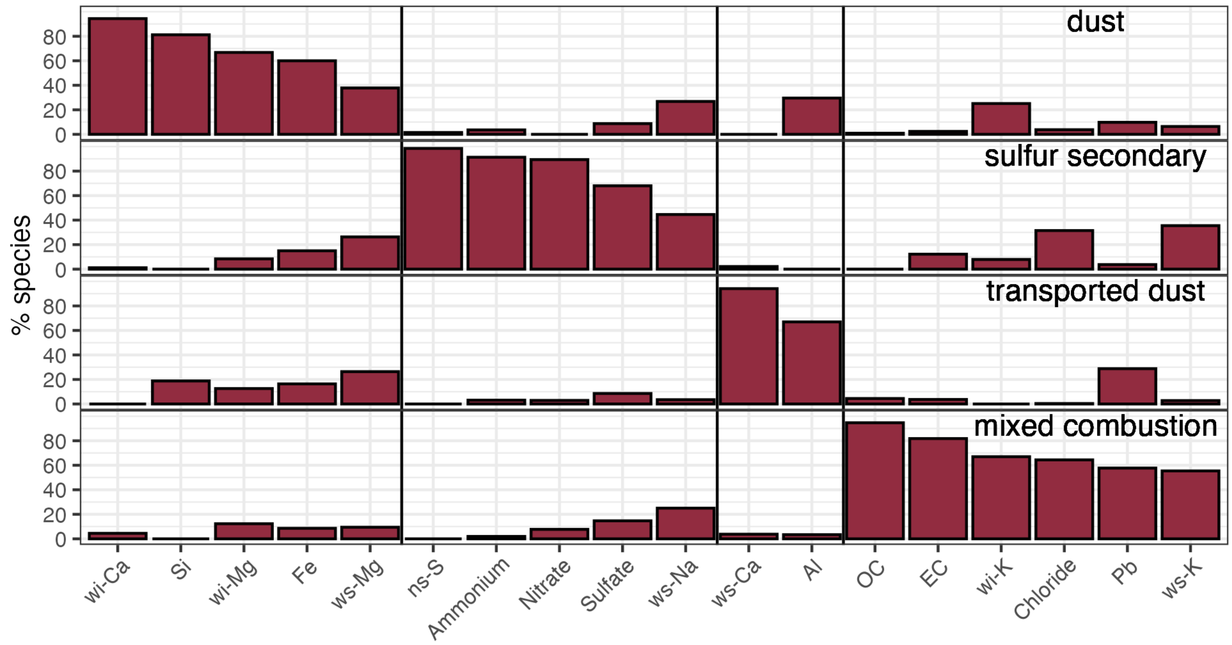


Figure 4.1. Source profiles identified by the four-factor PMF source apportionment model including outdoor and personal $PM_{2.5}$ exposure measurements. Ws- and wi- indicate the water soluble and water insoluble fractions, respectively. The lines separate the species with the highest loadings (% species) for each source.

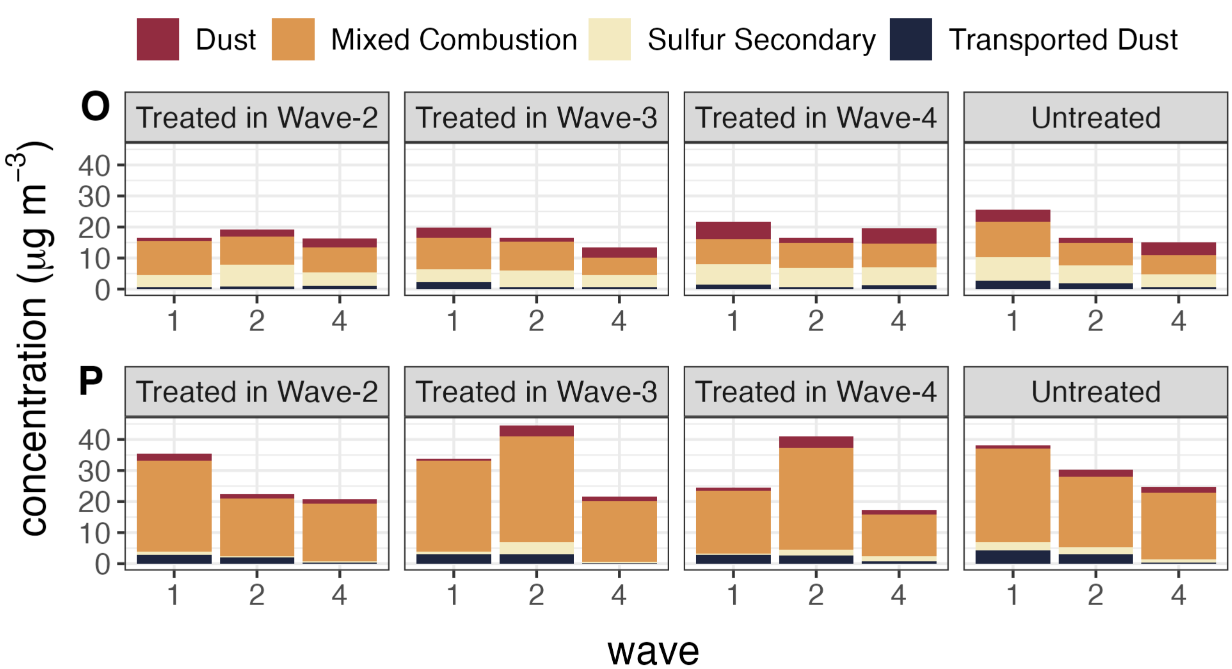


Figure 4.2. Average source contributions (μg m^{-3}) for the four sources identified by PMF for (O) outdoor and (P) personal exposure measurements by CHP treatment cohort and study wave.

We also identified a sulfur secondary source, which contained tracers of primary coal combustion (S), as well as secondary inorganic ions (ammonium, nitrate, sulfate) and is likely a mixture of primary and secondary sulfurous fuel combustion. The sulfur secondary source comprised nearly one third of the outdoor SSC (31.3% [28.1-34.4%]) and was a much higher percentage of outdoor than personal exposure (7.28% [5.28-9.29%]). Higher concentrations outdoors support our identification of this source as a secondary source given secondary species are formed outdoors. The percentage that the secondary sulfur source contributed to the SSC among personal exposure measurements was much lower than that of personal exposure in Beijing and Huairou (24.8% ± 32.6%).[117] The percentage the secondary sulfur source contributed to the SSC among outdoor samples collected in this study were similar to that report for outdoor PM samples collected in Gulou (33.0% ± 16.6%), a large city in the Yangtze River Delta region.[119]

The third source was found to be a transported dust (ws-Ca, Al), which likely results from long-range transport of dust that culminates in spring-time (e.g., March, April) dust storms.[120] The source contribution of the transported dust source was largest in Wave-1 (outdoor: 2.24 [1.96-2.52] $\mu g\ m^{-3}$; personal exposure: 3.85 [3.68-4.03] $\mu g\ m^{-3}$), second largest in Wave-2 (outdoor: 1.38 [1.21-1.54] $\mu g\ m^{-3}$; personal exposure: 2.87 [2.69-3.06] $\mu g\ m^{-3}$), and lowest in Wave-4 (outdoor: 0.81 [0.65-0.97] $\mu g\ m^{-3}$; personal exposure: 0.33 [0.24-0.43] $\mu g\ m^{-3}$). Additionally, transported dust was a larger percentage of SSC for personal exposure (10.6% [8.28-12.9%]) compared to outdoor measurements (7.28% [4.88-9.67%]). These differences in the contributions of the transported dust source between seasons corresponds with our field observations of dust storms occurring in the earlier study waves. While our measurements were not conducted in the spring, the settled particles from dust storms in the previous years were likely resuspended due to activity in the subsequent wintertime.

The final source identified was a mixed combustion source with tracers of coal (OC, wi-K, Chloride, Pb) and biomass (EC, ws-K) combustion. Some studies in the Beijing region have identified separate biomass and coal combustion sources, but oftentimes the source profiles have similar loadings for elements that can be tracers for both coal and biomass combustion.[121–123] A few studies have taken a more conservative approach and grouped the coal and biomass combustion sources together, as we chose to do in our study.[117,124] The mixed combustion source comprised 70.2% [67.2-73.3%] of personal exposure SSC and 46.9% [43.4-50.4%] of outdoor measurements. Higher percentages in personal exposure measurements compared to outdoor measurements further supports a coal- and biomass-based combustion source that we expected to be concentrated indoors where participants spend most of their time. The mixed combustion source being the largest identified contributing source to outdoor and personal exposure measurements also supports the idea that exposure to air pollution from solid fuel combustion is the main contributor to outdoor and personal air pollution exposure in our study.

We explored the spatial and temporal trends in the source contributions by visualizing them as a function of district and time (Appendix C3). For personal exposure measurements, one village was visited each day, district by district, so temporal trends are difficult to separate from village-level effects. We did see a noticeable increase in the transported dust source in Fangshan and Mentougou for both outdoor and personal exposure measurements, which suggests a spatial trend to this source. Visual differences by space or time for the other sources were not identified.

As a sensitivity analysis, we developed separate PMF models for outdoor and personal exposure measurements (Appendix C4). The models had some differences, with the three-factor outdoor models identifying dust, secondary inorganic ions, and mixed combustion, while the five-factor solution introduced a transported dust source, as well as an EC-and OC-dominated source. The three-factor, personal exposure model identified a transported dust, a dust, and a mixed combustion source while the five-factor solution additionally identified a secondary and EC/OC dominated source. Separating the sample types provided additional detail into the sources that were unique to outdoor and personal exposure $PM_{2.5}$ measurements, but it did not fundamentally change the identification of the key pollution source (mixed combustion).

Our study highlights the impacts of the selection of chemical analyses on the sources identified. With XRF, IC, and EC/OC, we did not identify a coal-specific source, but did conduct individual filter analyses, which greatly increased our sample size. More specific tracers of coal and biomass combustion would have likely been needed to separate the two sources, including tracers from destructive gas-chromatography mass-spectroscopy (GC-MS). However, given the short duration of sampling (24-hr for personal exposure) and the amount of organic matter required for GC-MS (~300 ug), compositing of our samples would have been needed. Compositing would have reduced our sample size from 1093 for personal exposure and 691 for outdoor measurements to 150 (one composite per village per season), greatly limiting our statistical power to observe significant treatment effects. It remains unclear how the tradeoff of sample size and source specificity (driven by choice of chemical analyses) would affect the estimation of treatment effects, but it is a potentially worthwhile area for future research. It should also be a key consideration when designing future studies that use similar source apportionment approaches to evaluate the impacts of household energy transitions.

4.3.3 Modeled impacts of the CHP on the mixed combustion source contribution

The average treatment effect for the treated (ATT) villages for outdoor and personal exposure to the mixed combustion source estimated by the ETWFE, DCF, and BCF models are shown in Figure 4.3. We observed no reduction in the mixed combustion source contribution to outdoor measurements estimated by any method. For personal exposure to the mixed combustion source, the BCF and DCF models indicated statistically imprecise reductions, and the ETWFE model showed a precise reduction (-7.75 [-13.4, -2.14] µg m^{-3}). The point estimates are nearly identical for the three ATT estimation methods. However, the confidence intervals tended to be smallest for the ETWFE model and largest for the DCF model. The mean-centered 95% credible interval for the BCF method was close to the 95% confidence interval of the ETWFE model for outdoor, but larger than the same interval for the personal exposure models. A joint F-test that all time-varying ATTs were equal (Appendix C5) did not indicate any differences in the time-varying treatment effects for outdoor (p-value = 0.60) or personal exposure models (p-value = 0.71).

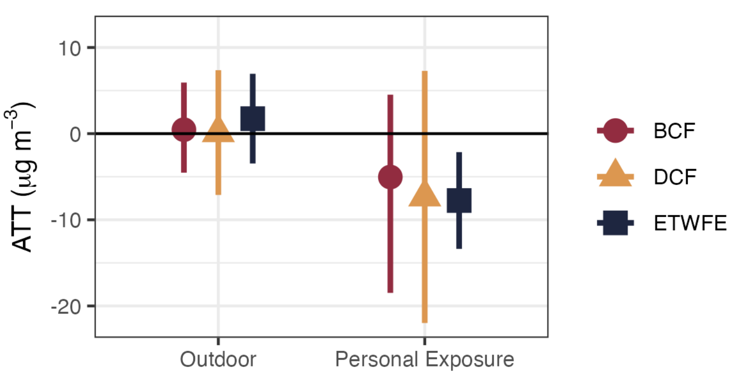


Figure 4.3. Average treatment effect for the treated (ATT; µg m^{-3}) estimated by the BCF, DCF, and ETWFE models for outdoor and personal exposure to the mixed combustion source contribution identified by the four-factor PMF source apportionment model.

The outdoor ETWFE model was adjusted for the village population and relative humidity as a spline with two degrees of freedom. The outdoor mixed combustion source concentration increased with village population and decreased with increasing relative humidity. The ETWFE personal exposure models were adjusted for participant smoking status, if the participant lived with a smoker but did not smoke themselves, and if the participant's household used wood as a solid fuel source. Wood use, smoking, and living with a smoker were positively associated with personal exposure to the mixed combustion source. The machine learning models were supplied with a larger number of covariates to estimate the propensity score including additional meteorological variables (temperature, wind speed, wind direction, boundary layer height) and self-reported area (square meters) of the home heated.

Our findings of the impacts of the CHP on the mixed combustion source did not vary with sensitivity analyses. A portion of the air pollution measured in our study was likely transported from nearby villages. Our estimation could be biased if a differential number of surrounding villages were enrolled in CHP between study waves, meaning we would expect the air pollution spillover to be lower if more surrounding villages were enrolled in the CHP. The number of villages within 20 km enrolled in the CHP remained similar between Wave-1 and Wave-2, suggesting that, to some extent, air pollution spillover is accounted for in our difference-in-difference study design (Appendix B10). We also conducted a sensitivity analysis where villages from Fangshan were excluded, which did not change the ATT estimation. No villages in Fangshan entered the CHP, so removing villages from Fangshan may improve the quality of the untreated group by only comparing villages that were in districts that were enrolled in the CHP.

Few studies have utilized source apportionment or factor analysis of air pollution samples to better understand the impacts of household energy interventions on air pollution. Piedrahita et al. 2017, conducted PMF to identify sources of organic matter and used the source contributions as an outcome to understand how various stove types impacted the source contributions in Northern Ghana.[125] Lai et al., 2019 apportioned the organic fraction of $PM_{2.5}$ to provide additional context on the indoor, outdoor, and personal $PM_{2.5}$ exposure sources during a cookstove intervention in Sichuan, China.[126] While both studies use source apportionment in the context of a household energy intervention, they use the source contributions to provide context to the main treatment effect estimate and not as a standalone outcome as was evaluated and presented here. Authors of both studies note the importance of source apportionment for providing additional insights in the main treatment effect on air pollution. Zhao et al., 2023 conducted a factor analysis for ambient $PM_{2.5}$ in a single village enrolled in our study, identified source tracers, compared those tracers pre- and post-intervention, and found a significant reduction in the tracers for coal.[97] While the results of this study disagree with out outdoor analysis, their study was limited by smaller sample size, lack of a control group, and no evaluation of source contributions. The Household Air Pollution Intervention Network (HAPIN) trial, a randomized control trial of LPG stoves, conducted a similar analysis of source contributions to personal PM exposure samples, as we did in our study, and found significant reductions in the source contribution from the targeted stove.[127] Taken together, previous studies and our work presented here support the idea that chemical analyses and source apportionment are useful tools for understanding the impacts of household energy interventions on air pollution. Specifically, our study demonstrates that, in the context of a real-world policy implementation, the using the source contributions of the targeted air pollution source as an outcome can provide unique insights on the air quality impacts of an air pollution intervention policy, additional to those gained by evaluating total $PM_{2.5}$ alone.

Our results that the CHP did not reduce outdoor concentrations of the mixed combustion source align with our findings that the CHP did not reduce outdoor air pollution. However, our results that the CHP reduced personal exposure to the coal-containing source provides additional context to our findings that the CHP did not reduce personal exposure to total $PM_{2.5}$. Personal exposure to total $PM_{2.5}$ reduced over time for both treated and untreated groups, so the reductions in personal $PM_{2.5}$ exposure could not be attributed to the CHP. The potential reason we identified reductions in the mixed combustion source but not total $PM_{2.5}$ is twofold. First, the source contributions are a more specific outcome of the hypothesized impacts of the policy. Second, isolating a single source limits the variability to that one source instead of compounding the variability from all sources of $PM_{2.5}$, which makes source contributions a more precise outcome to evaluate than total $PM_{2.5}$. Personal $PM_{2.5}$ exposure measurements are highly variable, so measuring outcomes that are more specific to the policy and more precise can improve the treatment effect estimation for personal exposure measurements.

Our study provides evidence that machine learning methods can accurately estimate the average and time-varying treatment effects as compared to the ETWFE model. However, the estimates were not as precise as those of the ETWFE models indicated by the larger confidence and credible intervals. A major limitation of the current machine learning implementation for our study design is the inability to include multiple, unpaired measurements from the unit of treatment assignment. To implement the village-level DiD design for CI machine learning methods, village level average concentrations of the mixed combustion source were used as the outcome, which reduced our sample size to ~140 villages across the three study waves. Further, since we did not take measurements in all 50 villages in Wave-1, some data were unavailable to be used in the machine learning analysis because we had to manually take the difference in means for villages between study waves. This reduction in sample size is likely the main contributor to the larger confidence and credible intervals observed for the machine learning methods. For smaller sample sizes, the machine learning methods may produce similar treatment effect estimations but are not as useful estimation methods compared to those currently implemented for DiD study designs similar to ours. However, these methods are likely still useful if the treatment effect is less variable or estimated with a large number of groups at the level of treatment assignment. Additional methodological improvements to the machine learning methods are needed to increase the precision or treatment effect estimation.

Our study is the first to evaluate the impacts of a real-world household energy transition policy on the source contribution to the targeted source. We conducted one of the largest, comprehensive measurement campaigns, collecting outdoor and personal exposure measurements in rural villages where the impact of the CHP had not been previously evaluated. Additionally, we performed individual filter chemical analyses, which enabled us to keep our large sample size in the treatment effect estimations. Finally, the comprehensive questionnaire gathered several key variables for our statistical models that let us adjust for the other sources likely contributing to the mixed combustion source.

Our study had several limitations. Comparing the trends in the untreated and treated in Wave-3 and Wave-4 groups for earlier years, we found evidence that the pre-trends differed for personal exposure to the mixed combustion source (Appendix C6). The pre-trend plots showed that personal exposure to the mixed combustion source decreased for untreated but increased for villages treated in Wave-3 and Wave-4 between Wave-1 and Wave-2. It is possible that the differing pre-trends reduced the treatment effects in these cohorts, and we underestimated the treatment effect. However, the difference in pre-trends for these cohorts suggest that the untreated villages were a poor control group for the later treated groups. It is unclear if these pre-trends applied to the villages treated in Wave-2, which was the largest treated cohort, since we did not take measurements for two years prior to Wave-2. Given the difference in pre-trends, our results provide some, and not definitive, evidence about the impacts of the CHP on personal exposure to the mixed combustion source. Additionally, we were not able to identify a coal-specific source to use as the outcome. However, we did adjust for wood use as well as personal and environmental tobacco smoke exposure, which were likely the other major contributing sources to the mixed combustion source. Third, the COVID-19 pandemic occurred during Wave-3, which may have affected how the policy was implemented in the later treated groups in ways that we were not able to quantify. This makes it challenging to separate policy impacts from the potential effects of differences in policy implementation due to COVID-19 among villages that were treated in later waves.

The Clean Heating Plan implemented by the Chinese government improved personal exposure to air pollution by reducing exposure to the coal-containing source contribution. We demonstrated how chemical analysis and source apportionment have the potential to provide additional insight into the impacts of a household energy transition on personal air pollution exposure. We found some evidence that personal exposure to the coal containing mixed combustion source reduced (-7.75 (-2.14, -13.4) $\mu g\ m^{-3}$ for treated groups as estimated by the ETWFE model. The treatment effect estimated by the BCF and DCF methods were similar to those estimated by the ETWFE model, but additional methodological development is needed to improve precision. By developing better outcomes and improving treatment effect estimation methods, we can better understand the impacts and better implement household energy transitions in the future

CHAPTER 5: CONCLUSIONS, RECCOMENDATIONS, AND FUTURE DIRECTION

This work presented several key findings related to the impact of the CHP on multiple outcomes. First, I showed that the secondary biomass-kang heating stove use did not change in related to the policy. Second, I found that the CHP had no impact on outdoor or personal exposure to $PM_{2.5}$ but did reduce seasonal indoor $PM_{2.5}$. Finally, I showed that the policy reduced personal exposure to the coal containing source contribution and not the outdoor contribution.

I demonstrated how data-driven methods can be used to develop important outcomes related to the policy. XGBoost machine learning models were deployed with low predictive error to predict stove use events after being trained on a subset of the data that had events hand-identified. Additionally, chemical analysis and source apportionment provided additional context to total $PM_{2.5}$ measurements that result in different conclusions about the impact of a household energy transition on air pollution outcomes.

Within each chapter, comparing the developed or measured outcomes to other relevant outcomes collected in the study exposed several key findings. First, I found that the objective heating stove temperature measures align with the impacts of the policy on self-reported measures of stove use (heating hours) but not the self-reported quantity of fuel used. Our indoor, outdoor, and personal $PM_{2.5}$ exposure measurements and mixed effect models provided additional context to why the post-intervention pollution measures remained high and potential reasons why the CHP had no impact on personal exposure. Lastly, I showed that the source apportionment approach provided additional insight into personal $PM_{2.5}$ exposure that the total $PM_{2.5}$ measurement alone.

I additionally showed how approaches to treatment effect estimation may have impacted our findings. Controlling for secular trends in the ETWFE models for personal $PM_{2.5}$ exposure were important. Had I not adjusted for difference in treatment timing or study wave, I would have found considerable impacts of the policy on personal exposure. In chapter 3, I show that machine learning methods can accurately estimate the treatment effect but are currently limited in their implementation when measurements are taken at the sub-treatment unit level and for a difference-in-difference approach.

Future evaluations of household energy interventions can learn from this work in several ways. A machine learning model trained on a subset of manually labeled stove use events can save considerable time manually labeling the entire dataset. Defining general event criteria can help with consistency of event identification and be used inform the predictor variables for the machine learning models. Understanding the relationship between heating stove usage and the outcome of interest can aid in defining the heating stove usage events. Labeling observations as event starts and ends as opposed to if the observation was part of an event likely makes for easier modeling since the criteria for if an observation is an event differs depending on where in the event the observation is.

Additionally, I provided evidence that self-reported heating outcomes may be a sufficient measure of stove use. Future studies may choose not to collect objective stove temperature measurements when the outcome is how often the stove is used. Collecting stove temperature measurements is a time and cost intensive process, so using self-reported measures of stove can allow research to allocate resources to other areas of the study such as enrolling additional participants or taking more air pollution measurements. The self-reported stove use data is likely most useful when there is consistent, daily usage of the stove compared to intermittent use.

Chapters 2 and 3 demonstrated that the impacts of a household energy transition can be distinct between outdoor, indoor, and personal $PM_{2.5}$ exposure. Future studies should collect measurement for all three locations to gain a comprehensive understanding of the policy's impact on air pollution. While the findings of the policy's impacts on outdoor and personal exposure to total $PM_{2.5}$ were the same, the impact did vary for outdoor and personal exposure to the mixed combustion source. Conducting impact analyses with measurements beyond outdoor sensor networks provides additional insights into a policy's impacts on personal exposure, which is a key driver for household energy transitions.

Finally, chemical analysis of air pollution measurements to identify the contribution of the targeted source can provide additional insight into the policy's impact. Future researchers should consider the impact of the choice of chemical analyses and sample size tradeoff. Less specific or lower cost chemical analysis can allow for a larger sample size but may not isolate the specific source targeted by the policy, which could be advantageous in settings where the additional sources can be adjusted for or there is limited crossover between the lower cost, less-specific chemical species. More specific or higher cost chemical analysis can provide higher granularity in the specific source but would likely reduce the sample size. Despite the lower sample size, the more specific chemical analyses could be advantageous in setting where there is crossover of chemical species between several sources and no way to account for levels or exposure to those sources in statistical analyses. Appropriate statistical power analyses could be conducted to assess the tradeoff between the chemical analysis approaches.

This work identified future research areas for evaluating household energy transitions. Determining if self-reported measures of fuel use align with a measure of event intensity (i.e., difference in temperature between event start and end) could strengthen the findings that self-reported measures of stove use align with the object stove use measurements. Further, how the stove use time-series aligns with high resolution continuous indoor $PM_{2.5}$ or participant location measured by GPS could be explored to better understand fluctuations in indoor or differences personal $PM_{2.5}$ exposure.

Attributing changes in personal $PM_{2.5}$ exposure to household energy transitions has been difficult. The changes in day-to-day personal $PM_{2.5}$ exposure may be too variable to identify impacts without additional information or with singular measurements. Additional information about personal exposure is needed to improve the confidence in identified impacts of a household energy transition on personal exposure. Future work to parse out the impacts of a policy on personal exposure could include collecting additional personal exposure measurements to increase statistical power, measuring personal exposure multiple times for the same participants to better understand seasonal variability, and incorporating high-resolution GPS and time activity patterns to control for differences in behaviors across study waves.

At present, the tradeoff between more specific chemical analyses and sample size for policy analysis is unclear. Future research that measures both less specific and more specific chemical analyses, conducts separate source apportionments with the less and more specific chemical components, and compares the impacts of the policy on the source contributions from the less and more specific source apportionment models would be useful for understanding the tradeoffs. Insight into the tradeoffs can inform the and budgeting and study design of future household energy transition evaluations

REFERENCES

(1) Bond, T. C.; Doherty, S. J.; Fahey, D. W.; Forster, P. M.; Berntsen, T.; DeAngelo, B. J.; Flanner, M. G.; Ghan, S.; Kärcher, B.; Koch, D.; Kinne, S.; Kondo, Y.; Quinn, P. K.; Sarofim, M. C.; Schultz, M. G.; Schulz, M.; Venkataraman, C.; Zhang, H.; Zhang, S.; Bellouin, N.; Guttikunda, S. K.; Hopke, P. K.; Jacobson, M. Z.; Kaiser, J. W.; Klimont, Z.; Lohmann, U.; Schwarz, J. P.; Shindell, D.; Storelvmo, T.; Warren, S. G.; Zender, C. S. Bounding the Role of Black Carbon in the Climate System: A Scientific Assessment. *Journal of Geophysical Research: Atmospheres* **2013**, *118* (11), 5380–5552. https://doi.org/10.1002/jgrd.50171.

(2) Dhimal, M.; Chirico, F.; Bista, B.; Sharma, S.; Chalise, B.; Dhimal, M. L.; Ilesanmi, O. S.; Trucillo, P.; Sofia, D. Impact of Air Pollution on Global Burden of Disease in 2019. *Processes* **2021**, *9* (10), 1719. https://doi.org/10.3390/pr9101719.

(3) WHO. *Household air pollution*. https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health (accessed 2024-12-09).

(4) Bonjour, S.; Adair-Rohani, H.; Wolf, J.; Bruce, N. G.; Mehta, S.; Prüss-Ustün, A.; Lahiff, M.; Rehfuess, E. A.; Mishra, V.; Smith, K. R. Solid Fuel Use for Household Cooking: Country and Regional Estimates for 1980–2010. *Environ Health Perspect* **2013**, *121* (7), 784–790. https://doi.org/10.1289/ehp.1205987.

(5) Bowe, B.; Xie, Y.; Li, T.; Yan, Y.; Xian, H.; Al-Aly, Z. The 2016 Global and National Burden of Diabetes Mellitus Attributable to PM2·5 Air Pollution. *The Lancet Planetary Health* **2018**, *2* (7), e301–e312. https://doi.org/10.1016/S2542-5196(18)30140-2.

(6) Li, H.; Zhou, D.; Zhang, Q.; Feng, C.; Zheng, W.; He, K.; Lan, Y. Vanadium Exposure-Induced Neurobehavioral Alterations among Chinese Workers. *NeuroToxicology* **2013**, *36*, 49–54. https://doi.org/10.1016/j.neuro.2013.02.008.

(7) Baumgartner, J.; Zhang, Y.; Schauer, J. J.; Ezzati, M.; Patz, J. A.; Bautista, L. E. Household Air Pollution and Children's Blood Pressure. *Epidemiology* **2012**, *23* (4), 641–642. https://doi.org/10.1097/EDE.0b013e3182593fa9.

(8) Jin, W.; Su, S.; Wang, B.; Zhu, X.; Chen, Y.; Shen, G.; Liu, J.; Cheng, H.; Wang, X.; Wu, S.; Zeng, E.; Xing, B.; Tao, S. Properties and Cellular Effects of Particulate Matter from Direct Emissions and Ambient Sources. *Journal of Environmental Science and Health, Part A* **2016**, *51* (12), 1075–1083. https://doi.org/10.1080/10934529.2016.1198632.

(9) Pagan, I.; Costa, D. L.; McGee, J. K.; Richards, J. H.; Dye, J. A.; Dykstra, M. J. Metals Mimic Airway Epithelial Injury Induced by in Vitro Exposure to Utah Valley Ambient Particulate Matter Extracts. *Journal of Toxicology and Environmental Health, Part A* **2003**, *66* (12), 1087–1112. https://doi.org/10.1080/15287390390213908.

(10) Verma, M.; Pervez, S.; Majumdar, D.; Chakrabarty, R.; Pervez, Y. F. Emission Estimation of Aromatic and Halogenated VOCs from Household Solid Fuel Burning Practices. *Int. J. Environ. Sci. Technol.* **2019**, *16* (6), 2683–2692. https://doi.org/10.1007/s13762-018-1920-7.

(11) Ozgen, S.; Caserini, S. Methane Emissions from Small Residential Wood Combustion Appliances: Experimental Emission Factors and Warming Potential. *Atmospheric Environment* **2018**, *189*, 164–173. https://doi.org/10.1016/j.atmosenv.2018.07.006.

(12) Wei, W.; Zhang, W.; Hu, D.; Ou, L.; Tong, Y.; Shen, G.; Shen, H.; Wang, X. Emissions of Carbon Monoxide and Carbon Dioxide from Uncompressed and Pelletized Biomass Fuel Burning in Typical Household Stoves in China. *Atmospheric Environment* **2012**, *56*, 136–142. https://doi.org/10.1016/j.atmosenv.2012.03.060.
(13) Jin, Y.; Andersson, H.; Zhang, S. Air Pollution Control Policies in China: A Retrospective and Prospects. *International Journal of Environmental Research and Public Health* **2016**, *13* (12), 1219. https://doi.org/10.3390/ijerph13121219.
(14) Feng, L.; Liao, W. Legislation, Plans, and Policies for Prevention and Control of Air Pollution in China: Achievements, Challenges, and Improvements. *Journal of Cleaner Production* **2016**, *112*, 1549–1558. https://doi.org/10.1016/j.jclepro.2015.08.013.
(15) Shen, G.; Ru, M.; Du, W.; Zhu, X.; Zhong, Q.; Chen, Y.; Shen, H.; Yun, X.; Meng, W.; Liu, J.; Cheng, H.; Hu, J.; Guan, D.; Tao, S. Impacts of Air Pollutants from Rural Chinese Households under the Rapid Residential Energy Transition. *Nat Commun* **2019**, *10* (1), 3405. https://doi.org/10.1038/s41467-019-11453-w.
(16) Group, G. M. W. Burden of Disease Attributable to Coal-Burning and Other Major Sources of Air Pollution in China. *Special report* **2016**, *20*.
(17) Dispersed Coal Management Research Group. China Dispersed Coal Governance Report, 2023.
(18) Quansah, R.; Semple, S.; Ochieng, C. A.; Juvekar, S.; Armah, F. A.; Luginaah, I.; Emina, J. 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* **2017**, *103*, 73–90.
(19) Dockery, D. W.; Rich, D. Q.; Goodman, P. G.; Clancy, L.; Ohman-Strickland, P.; George, P.; Kotlov, T. Effect of Air Pollution Control on Mortality and Hospital Admissions in Ireland. *Research Report (Health Effects Institute)* **2013**, No. 176, 3–109.
(20) Johnston, F. H.; Hanigan, I. C.; Henderson, S. B.; Morgan, G. G. Evaluation of Interventions to Reduce Air Pollution from Biomass Smoke on Mortality in Launceston, Australia: Retrospective Analysis of Daily Mortality, 1994-2007. *Bmj* **2013**, *346*.
(21) Scott, A. J.; Scarrott, C. Impacts of Residential Heating Intervention Measures on Air Quality and Progress towards Targets in Christchurch and Timaru, New Zealand. *Atmospheric environment* **2011**, *45* (17), 2972–2980.
(22) Meng, W.; Zhu, L.; Liang, Z.; Xu, H.; Zhang, W.; Li, J.; Zhang, Y.; Luo, Z.; Shen, G.; Shen, H.; Chen, Y.; Cheng, H.; Ma, J.; Tao, S. Significant but Inequitable Cost-Effective Benefits of a Clean Heating Campaign in Northern China. *Environ. Sci. Technol.* **2023**, *57* (23), 8467–8475. https://doi.org/10.1021/acs.est.2c07492.
(23) Zhang, X.; Jin, Y.; Dai, H.; Xie, Y.; Zhang, S. Health and Economic Benefits of Cleaner Residential Heating in the Beijing–Tianjin–Hebei Region in China. *Energy Policy* **2019**, *127*, 165–178. https://doi.org/10.1016/j.enpol.2018.12.008.
(24) Zhang, J.; Wang, W.; Gao, L.; Deng, Z.; Tian, Y. Can the Coal-to-Gas/Electricity Policy Improve Air Quality in the Beijing–Tianjin–Hebei Region?—Empirical Analysis Based on the PSM-DID. *Atmosphere* **2022**, *13* (6), 879. https://doi.org/10.3390/atmos13060879.
(25) Gehring, U.; Wijga, A. H.; Brauer, M.; Fischer, P.; de Jongste, J. C.; Kerkhof, M.; Oldenwening, M.; Smit, H. A.; Brunekreef, B. Traffic-Related Air Pollution and the Development of Asthma and Allergies during the First 8 Years of Life. *Am J Respir Crit Care Med* **2010**, *181* (6), 596–603. https://doi.org/10.1164/rccm.200906-0858OC.

(26) Yang, A.; Janssen, N. A.; Brunekreef, B.; Cassee, F. R.; Hoek, G.; Gehring, U. Children's Respiratory Health and Oxidative Potential of PM2. 5: The PIAMA Birth Cohort Study. *Occup Environ Med* **2016**, *73* (3), 154–160.
(27) Manibog, F. R. Improved Cooking Stoves in Developing Countries: Problems and Opportunities. *Annual Review of Environment and Resources* **1984**, *9* (Volume 9, 1984), 199–227. https://doi.org/10.1146/annurev.eg.09.110184.001215.
(28) Kshirsagar, M. P.; Kalamkar, V. R. A Comprehensive Review on Biomass Cookstoves and a Systematic Approach for Modern Cookstove Design. *Renewable and Sustainable Energy Reviews* **2014**, *30*, 580–603. https://doi.org/10.1016/j.rser.2013.10.039.
(29) Rehfuess, E. A.; Puzzolo, E.; Stanistreet, D.; Pope, D.; Bruce, N. G. Enablers and Barriers to Large-Scale Uptake of Improved Solid Fuel Stoves: A Systematic Review. *Environmental Health Perspectives* **2014**, *122* (2), 120–130. https://doi.org/10.1289/ehp.1306639.
(30) Ruiz-Mercado, I.; Masera, O. Patterns of Stove Use in the Context of Fuel–Device Stacking: Rationale and Implications. *EcoHealth* **2015**, *12* (1), 42–56. https://doi.org/10.1007/s10393-015-1009-4.
(31) Zhu, X.; Yun, X.; Meng, W.; Xu, H.; Du, W.; Shen, G.; Cheng, H.; Ma, J.; Tao, S. Stacked Use and Transition Trends of Rural Household Energy in Mainland China. *Environ. Sci. Technol.* **2019**, *53* (1), 521–529. https://doi.org/10.1021/acs.est.8b04280.
(32) Baumgartner, J.; Clark, S.; Carter, E.; Lai, A.; Zhang, Y.; Shan, M.; Schauer, J. J.; Yang, X. Effectiveness of a Household Energy Package in Improving Indoor Air Quality and Reducing Personal Exposures in Rural China. *Environ. Sci. Technol.* **2019**, *53* (15), 9306–9316. https://doi.org/10.1021/acs.est.9b02061.
(33) Burns, J.; Boogaard, H.; Polus, S.; Pfadenhauer, L. M.; Rohwer, A. C.; van Erp, A. M.; Turley, R.; Rehfuess, E. A. Interventions to Reduce Ambient Air Pollution and Their Effects on Health: An Abridged Cochrane Systematic Review. *Environment International* **2020**, *135*, 105400. https://doi.org/10.1016/j.envint.2019.105400.
(34) Allen, R. W.; Leckie, S.; Millar, G.; Brauer, M. The Impact of Wood Stove Technology Upgrades on Indoor Residential Air Quality. *Atmospheric Environment* **2009**, *43* (37), 5908–5915. https://doi.org/10.1016/j.atmosenv.2009.08.016.
(35) Johnson, M. A.; Steenland, K.; Piedrahita, R.; Clark, M. L.; Pillarisetti, A.; Balakrishnan, K.; Peel, J. L.; Naeher, L. P.; Liao, J.; Wilson, D.; Sarnat, J.; Underhill, L. J.; Burrowes, V.; McCracken, J. P.; Rosa, G.; Rosenthal, J.; Sambandam, S.; de Leon, O.; Kirby, M. A.; Kearns, K.; Checkley, W.; Clasen, T.; HAPIN Investigators. Air Pollutant Exposure and Stove Use Assessment Methods for the Household Air Pollution Intervention Network (HAPIN) Trial. *Environmental Health Perspectives* **2020**, *128* (4), 047009. https://doi.org/10.1289/EHP6422.
(36) Piedrahita, R.; Johnson, M.; Bilsback, K. R.; L'Orange, C.; Kodros, J. K.; Eilenberg, S. R.; Naluwagga, A.; Shan, M.; Sambandam, S.; Clark, M.; Pierce, J. R.; Balakrishnan, K.; Robinson, A. L.; Volckens, J. Comparing Regional Stove-Usage Patterns and Using Those Patterns to Model Indoor Air Quality Impacts. *Indoor Air* **2020**, *30* (3), 521–533. https://doi.org/10.1111/ina.12645.
(37) Clark, S.; Carter, E.; Shan, M.; Ni, K.; Niu, H.; Tseng, J. T. W.; Pattanayak, S. K.; Jeuland, M.; Schauer, J. J.; Ezzati, M.; C Wiedinmyer; Yang, X.; Baumgartner, J. Adoption and Use of a Semi-Gasifier Cooking and Water Heating Stove and Fuel Intervention in the Tibetan

Plateau, China. *Environ. Res. Lett.* **2017**, *12* (7), 075004. https://doi.org/10.1088/1748-9326/aa751e.
(38) Young, B. N.; Peel, J. L.; Benka-Coker, M. L.; Rajkumar, S.; Walker, E. S.; Brook, R. D.; Nelson, T. L.; Volckens, J.; L'Orange, C.; Good, N.; Quinn, C.; Keller, J. P.; Weller, Z. D.; Africano, S.; Osorto Pinel, A. B.; Clark, M. L. Study Protocol for a Stepped-Wedge Randomized Cookstove Intervention in Rural Honduras: Household Air Pollution and Cardiometabolic Health. *BMC Public Health* **2019**, *19* (1), 903. https://doi.org/10.1186/s12889-019-7214-2.
(39) Ruiz-Mercado, I.; Canuz, E.; Walker, J. L.; Smith, K. R. Quantitative Metrics of Stove Adoption Using Stove Use Monitors (SUMs). *Biomass and Bioenergy* **2013**, *57*, 136–148. https://doi.org/10.1016/j.biombioe.2013.07.002.
(40) Snider, G.; Carter, E.; Clark, S.; Tseng, J. (Tzu W.; Yang, X.; Ezzati, M.; Schauer, J. J.; Wiedinmyer, C.; Baumgartner, J. Impacts of Stove Use Patterns and Outdoor Air Quality on Household Air Pollution and Cardiovascular Mortality in Southwestern China. *Environment International* **2018**, *117*, 116–124. https://doi.org/10.1016/j.envint.2018.04.048.
(41) Northcross, A.; Shupler, M.; Alexander, D.; Olamijulo, J.; Ibigbami, T.; Ana, G.; Ojengbede, O.; Olopade, C. O. Sustained Usage of Bioethanol Cookstoves Shown in an Urban Nigerian City via New SUMs Algorithm. *Energy for Sustainable Development* **2016**, *35*, 35–40. https://doi.org/10.1016/j.esd.2016.05.003.
(42) Piedrahita, R.; Dickinson, K. L.; Kanyomse, E.; Coffey, E.; Alirigia, R.; Hagar, Y.; Rivera, I.; Oduro, A.; Dukic, V.; Wiedinmyer, C.; Hannigan, M. Assessment of Cookstove Stacking in Northern Ghana Using Surveys and Stove Use Monitors. *Energy for Sustainable Development* **2016**, *34*, 67–76. https://doi.org/10.1016/j.esd.2016.07.007.
(43) Wilson, D. L.; Monga, M.; Saksena, A.; Kumar, A.; Gadgil, A. Effects of USB Port Access on Advanced Cookstove Adoption. *Development Engineering* **2018**, *3*, 209–217. https://doi.org/10.1016/j.deveng.2018.08.001.
(44) Perros, T.; Büttner, P.; Leary, J.; Parikh, P. Pay-as-You-Go LPG: A Mixed-Methods Pilot Study in Urban Rwanda. *Energy for Sustainable Development* **2021**, *65*, 117–129. https://doi.org/10.1016/j.esd.2021.10.003.
(45) Li, X.; Baumgartner, J.; Barrington-Leigh, C.; Harper, S.; Robinson, B.; Shen, G.; Sternbach, T.; Tao, S.; Zhang, X.; Zhang, Y.; Carter, E. Socioeconomic and Demographic Associations with Wintertime Air Pollution Exposures at Household, Community, and District Scales in Rural Beijing, China. *Environ. Sci. Technol.* **2022**, *56* (12), 8308–8318. https://doi.org/10.1021/acs.est.1c07402.
(46) *The ERA5 global reanalysis - Hersbach - 2020 - Quarterly Journal of the Royal Meteorological Society - Wiley Online Library*. https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3803 (accessed 2021-09-09).
(47) Li, C.; Ye, K.; Mawusi, S.; Zhang, W.; Xu, Y.; Xu, J.; Zhou, W.; Li, J.; Jiao, M.; Shrestha, P.; Pang, R.; Hussein, R.; Xue, C.; Liu, G. A 24-h Real-Time Emissions Assessment of 41 Uncontrolled Household Raw Coal Combustion Stoves in Four Provinces of Northern China. *Atmospheric Environment* **2020**, *235*, 117588. https://doi.org/10.1016/j.atmosenv.2020.117588.
(48) Kapoor, R.; Pillarisetti, A.; Coyle, J.; Croshere, S.; Paré, M.; Benkhay, H.; Wilson, D. Geocene/Trainset. https://github.com/Geocene/trainset.

(49) Chen, T.; Guestrin, C. Xgboost: A Scalable Tree Boosting System. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*; 2016; pp 785–794.
(50) Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; Bing, G. Learning from Class-Imbalanced Data: Review of Methods and Applications. *Expert systems with applications* **2017**, *73*, 220–239.
(51) Zou, Q.; Xie, S.; Lin, Z.; Wu, M.; Ju, Y. Finding the Best Classification Threshold in Imbalanced Classification. *Big Data Research* **2016**, *5*, 2–8. https://doi.org/10.1016/j.bdr.2015.12.001.
(52) Lechner, M. *The Estimation of Causal Effects by Difference-in-Difference Methods*; Now Hanover, MA, 2011; Vol. 4.
(53) Wooldridge, J. M. Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators. Rochester, NY August 17, 2021. https://doi.org/10.2139/ssrn.3906345.
(54) Pillarisetti, A.; Vaswani, M.; Jack, D.; Balakrishnan, K.; Bates, M. N.; Arora, N. K.; Smith, K. R. Patterns of Stove Usage after Introduction of an Advanced Cookstove: The Long-Term Application of Household Sensors. *Environ. Sci. Technol.* **2014**, *48* (24), 14525–14533. https://doi.org/10.1021/es504624c.
(55) Baumgartner, J.; Harper, S.; Barrington-Leigh, C.; Brehmer, C.; Carter, E.; Li, X.; Robinson, B.; Shen, G.; Sternbach, T.; Tao, S.; Xue, K.; Yuan, W.; Zhang, X.; Zhang, Y. *How Do Household Energy Transitions Work?*
(56) Fuller, R.; Landrigan, P. J.; Balakrishnan, K.; Bathan, G.; Bose-O'Reilly, S.; Brauer, M.; Caravanos, J.; Chiles, T.; Cohen, A.; Corra, L.; Cropper, M.; Ferraro, G.; Hanna, J.; Hanrahan, D.; Hu, H.; Hunter, D.; Janata, G.; Kupka, R.; Lanphear, B.; Lichtveld, M.; Martin, K.; Mustapha, A.; Sanchez-Triana, E.; Sandilya, K.; Schaefli, L.; Shaw, J.; Seddon, J.; Suk, W.; Téllez-Rojo, M. M.; Yan, C. Pollution and Health: A Progress Update. *The Lancet Planetary Health* **2022**, *6* (6), e535–e547. https://doi.org/10.1016/S2542-5196(22)00090-0.
(57) WHO. *Household air pollution*. https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health (accessed 2024-05-15).
(58) Yun, X.; Shen, G.; Shen, H.; Meng, W.; Chen, Y.; Xu, H.; Ren, Y.; Zhong, Q.; Du, W.; Ma, J.; Cheng, H.; Wang, X.; Liu, J.; Wang, X.; Li, B.; Hu, J.; Wan, Y.; Tao, S. Residential Solid Fuel Emissions Contribute Significantly to Air Pollution and Associated Health Impacts in China. *Science Advances* **2020**, *6* (44), eaba7621. https://doi.org/10.1126/sciadv.aba7621.
(59) Chafe, Z. A.; Brauer, M.; Klimont, Z.; Van Dingenen, R.; Mehta, S.; Rao, S.; Riahi, K.; Dentener, F.; Smith, K. R. Household Cooking with Solid Fuels Contributes to Ambient PM2.5 Air Pollution and the Burden of Disease. *Environmental Health Perspectives* **2014**, *122* (12), 1314–1320. https://doi.org/10.1289/ehp.1206340.
(60) Phogole, B.; Kelso, C.; Langerman, K. E. The Effectiveness of Household Energy Transition Interventions in a Coal-Using Community on the South African Highveld. *Energy for Sustainable Development* **2022**, *71*, 1–12. https://doi.org/10.1016/j.esd.2022.09.006.
(61) Guta, D.; Baumgartner, J.; Jack, D.; Carter, E.; Shen, G.; Orgill-Meyer, J.; Rosenthal, J.; Dickinson, K.; Bailis, R.; Masuda, Y.; Zerriffi, H. A Systematic Review of Household Energy Transition in Low and Middle Income Countries. *Energy Research & Social Science* **2022**, *86*, 102463. https://doi.org/10.1016/j.erss.2021.102463.

(62) Liao, J.; Kirby, M. A.; Pillarisetti, A.; Piedrahita, R.; Balakrishnan, K.; Sambandam, S.; Mukhopadhyay, K.; Ye, W.; Rosa, G.; Majorin, F.; Dusabimana, E.; Ndagijimana, F.; McCracken, J. P.; Mollinedo, E.; de Leon, O.; Díaz-Artiga, A.; Thompson, L. M.; Kearns, K. A.; Naeher, L.; Rosenthal, J.; Clark, M. L.; Steenland, K.; Waller, L. A.; Checkley, W.; Peel, J. L.; Clasen, T.; Johnson, M.; Aravindalochanan, V.; Bankundiye, G.; Barr, D. B.; Bussalleu, A.; Canuz, E.; Castañaza, A.; Chen, Y.; Chiang, M.; Craik, R.; Davila-Roman, V. G.; de las Fuentes, L.; Elon, L.; Espinoza, J. G.; Garg, S.; Hamid, S.; Hartinger, S.; Harvey, S. A.; Hengstermann, M.; Hennessee, I.; Herrera, P. M.; Hossen, S.; Howards, P. P.; Jaacks, L.; Jabbarzadeh, S.; Lenzen, P.; Lovvorn, A. E.; Mbabazi, J.; McCollum, E.; Meyers, R.; Moulton, L.; Mukeshimana, A.; Mutariyani, B.; Natesan, D.; Nizam, A.; Ntivuguruzwa, J. de D.; Papageorghiou, A.; Puttaswamy, N.; Puzzolo, E.; Quinn, A.; Rajamani, K. D.; Ramakrishnan, U.; Ramasami, R.; Ramirez, A.; Ryan, P. B.; Saidam, S.; Sarnat, J. A.; Simkovich, S.; Sinharoy, S. S.; Smith, K. R.; Swearing, D.; Thangavel, G.; Toenjes, A.; Valdes, V.; Williams, K. N.; Ye, W.; Young, B. N. LPG Stove and Fuel Intervention among Pregnant Women Reduce Fine Particle Air Pollution Exposures in Three Countries: Pilot Results from the HAPIN Trial. *Environmental Pollution* **2021**, *291*, 118198. https://doi.org/10.1016/j.envpol.2021.118198.

(63) Bruce, N.; Pope, D.; Rehfuess, E.; Balakrishnan, K.; Adair-Rohani, H.; Dora, C. WHO Indoor Air Quality Guidelines on Household Fuel Combustion: Strategy Implications of New Evidence on Interventions and Exposure–Risk Functions. *Atmospheric Environment* **2015**, *106*, 451–457. https://doi.org/10.1016/j.atmosenv.2014.08.064.

(64) Zhou, M.; Liu, H.; Peng, L.; Qin, Y.; Chen, D.; Zhang, L.; Mauzerall, D. L. Environmental Benefits and Household Costs of Clean Heating Options in Northern China. *Nat Sustain* **2022**, *5* (4), 329–338. https://doi.org/10.1038/s41893-021-00837-w.

(65) Zhang, Y.; Li, W.; Wu, F. Does Energy Transition Improve Air Quality? Evidence Derived from China's Winter Clean Heating Pilot (WCHP) Project. *Energy* **2020**, *206*, 118130. https://doi.org/10.1016/j.energy.2020.118130.

(66) National Development and Reform Commission. *Winter cleaning and heating plan for the northern region (2017–2021)*. https://householdenergypolicies.org/policy/191 (accessed 2023-10-26).

(67) Zhang, Y.; Li, W.; Wu, F. Does Energy Transition Improve Air Quality? Evidence Derived from China's Winter Clean Heating Pilot (WCHP) Project. *Energy* **2020**, *206*, 118130. https://doi.org/10.1016/j.energy.2020.118130.

(68) Bilsback, K. R.; Baumgartner, J.; Cheeseman, M.; Ford, B.; Kodros, J. K.; Li, X.; Ramnarine, E.; Tao, S.; Zhang, Y.; Carter, E. Estimated Aerosol Health and Radiative Effects of the Residential Coal Ban in the Beijing-Tianjin-Hebei Region of China. *Aerosol and Air Quality Research* **2020**, *20* (11), 2332–2346.

(69) Duan, X.; Jiang, Y.; Wang, B.; Zhao, X.; Shen, G.; Cao, S.; Huang, N.; Qian, Y.; Chen, Y.; Wang, L. Household Fuel Use for Cooking and Heating in China: Results from the First Chinese Environmental Exposure-Related Human Activity Patterns Survey (CEERHAPS). *Applied Energy* **2014**, *136*, 692–703. https://doi.org/10.1016/j.apenergy.2014.09.066.

(70) Li, X.; Baumgartner, J.; Barrington-Leigh, C.; Harper, S.; Robinson, B.; Shen, G.; Sternbach, T.; Tao, S.; Zhang, X.; Zhang, Y.; Carter, E. Socioeconomic and Demographic Associations with Wintertime Air Pollution Exposures at Household, Community, and District Scales in Rural Beijing, China. *Environ. Sci. Technol.* **2022**, *56* (12), 8308–8318. https://doi.org/10.1021/acs.est.1c07402.

(71) Li, X.; Baumgartner, J.; Harper, S.; Zhang, X.; Sternbach, T.; Barrington-Leigh, C.; Brehmer, C.; Robinson, B.; Shen, G.; Zhang, Y.; Tao, S.; Carter, E. Field Measurements of Indoor and Community Air Quality in Rural Beijing before, during, and after the COVID-19 Lockdown. *Indoor Air* **2022**, *32* (8), e13095. https://doi.org/10.1111/ina.13095.
(72) Shen, H.; Shen, G.; Chen, Y.; Russell, A. G.; Hu, Y.; Duan, X.; Meng, W.; Xu, Y.; Yun, X.; Lyu, B.; Zhao, S.; Hakami, A.; Guo, J.; Tao, S.; Smith, K. R. Increased Air Pollution Exposure among the Chinese Population during the National Quarantine in 2020. *Nat Hum Behav* **2021**, *5* (2), 239–246. https://doi.org/10.1038/s41562-020-01018-z.
(73) Sternbach, T. J.; Harper, S.; Li, X.; Zhang, X.; Carter, E.; Zhang, Y.; Shen, G.; Fan, Z.; Zhao, L.; Tao, S.; Baumgartner, J. Effects of Indoor and Outdoor Temperatures on Blood Pressure and Central Hemodynamics in a Wintertime Longitudinal Study of Chinese Adults. *Journal of Hypertension* **2022**, *40* (10), 1950. https://doi.org/10.1097/HJH.0000000000003198.
(74) Khandker, S.; B. Koolwal, G.; Samad, H. *Handbook on Impact Evaluation: Quantitative Methods and Practices*; The World Bank, 2009. https://doi.org/10.1596/978-0-8213-8028-4.
(75) Underhill, L. J.; Dols, W. S.; Lee, S. K.; Fabian, M. P.; Levy, J. I. Quantifying the Impact of Housing Interventions on Indoor Air Quality and Energy Consumption Using Coupled Simulation Models. *J Expo Sci Environ Epidemiol* **2020**, *30* (3), 436–447. https://doi.org/10.1038/s41370-019-0197-3.
(76) Wang, Q.; Fan, J.; Kwan, M.-P.; Zhou, K.; Shen, G.; Li, N.; Wu, B.; Lin, J. Examining Energy Inequality under the Rapid Residential Energy Transition in China through Household Surveys. *Nat Energy* **2023**, *8* (3), 251–263. https://doi.org/10.1038/s41560-023-01193-z.
(77) Rosa, G.; Majorin, F.; Boisson, S.; Barstow, C.; Johnson, M.; Kirby, M.; Ngabo, F.; Thomas, E.; Clasen, T. Assessing the Impact of Water Filters and Improved Cook Stoves on Drinking Water Quality and Household Air Pollution: A Randomised Controlled Trial in Rwanda. *PloS one* **2014**, *9* (3), e91011.
(78) Romieu, I.; Riojas-Rodriguez, H.; Marrón-Mares, A. T.; Schilmann, A.; Perez-Padilla, R.; Masera, O. Improved Biomass Stove Intervention in Rural Mexico: Impact on the Respiratory Health of Women. *American journal of respiratory and critical care medicine* **2009**, *180* (7), 649–656.
(79) Baumgartner, J.; Clark, S.; Carter, E.; Lai, A.; Zhang, Y.; Shan, M.; Schauer, J. J.; Yang, X. Effectiveness of a Household Energy Package in Improving Indoor Air Quality and Reducing Personal Exposures in Rural China. *Environ. Sci. Technol.* **2019**, *53* (15), 9306–9316. https://doi.org/10.1021/acs.est.9b02061.
(80) Aung, T. W.; Jain, G.; Sethuraman, K.; Baumgartner, J.; Reynolds, C.; Grieshop, A. P.; Marshall, J. D.; Brauer, M. Health and Climate-Relevant Pollutant Concentrations from a Carbon-Finance Approved Cookstove Intervention in Rural India. *Environ. Sci. Technol.* **2016**, *50* (13), 7228–7238. https://doi.org/10.1021/acs.est.5b06208.
(81) Vodonos, A.; Awad, Y. A.; Schwartz, J. The Concentration-Response between Long-Term PM2.5 Exposure and Mortality; A Meta-Regression Approach. *Environmental Research* **2018**, *166*, 677–689. https://doi.org/10.1016/j.envres.2018.06.021.
(82) Clark, M. L.; Peel, J. L.; Burch, J. B.; Nelson, T. L.; Robinson, M. M.; Conway, S.; Bachand, A. M.; Reynolds, S. J. Impact of Improved Cookstoves on Indoor Air Pollution and Adverse Health Effects among Honduran Women. *International Journal of*

Environmental Health Research **2009**, *19* (5), 357–368. https://doi.org/10.1080/09603120902842705.
(83) Zhang, J.; Wang, W.; Gao, L.; Deng, Z.; Tian, Y. Can the Coal-to-Gas/Electricity Policy Improve Air Quality in the Beijing–Tianjin–Hebei Region?—Empirical Analysis Based on the PSM-DID. *Atmosphere* **2022**, *13* (6), 879. https://doi.org/10.3390/atmos13060879.
(84) Li, X.; Tryner, J.; Young, B. N.; Hernandez Ramirez, L.; Phillips, M.; WeMott, S.; Erlandson, G.; Kuiper, G.; Dean, D.; Martinez, N.; Sanpedro, L.; Magzamen, S.; Volckens, J. Application and Validation of a Wearable Monitor for Assessing Time- and Location-Resolved Exposures to Particulate Matter in California's Central Valley. *Aerosol Science and Technology* **2024**, *0* (0), 1–17. https://doi.org/10.1080/02786826.2024.2415481.
(85) Tryner, J.; Quinn, C.; Molina Rueda, E.; Andales, M. J.; L'Orange, C.; Mehaffy, J.; Carter, E.; Volckens, J. AirPen: A Wearable Monitor for Characterizing Exposures to Particulate Matter and Volatile Organic Compounds. *Environ. Sci. Technol.* **2023**, acs.est.3c02238. https://doi.org/10.1021/acs.est.3c02238.
(86) Bu, X.; Xie, Z.; Liu, J.; Wei, L.; Wang, X.; Chen, M.; Ren, H. Global PM2.5-Attributable Health Burden from 1990 to 2017: Estimates from the Global Burden of Disease Study 2017. *Environmental Research* **2021**, *197*, 111123. https://doi.org/10.1016/j.envres.2021.111123.
(87) Gakidou, E.; Afshin, A.; Abajobir, A. A.; Abate, K. H.; Abbafati, C.; Abbas, K. M.; Abd-Allah, F.; Abdulle, A. M.; Abera, S. F.; Aboyans, V. Global, Regional, and National Comparative Risk Assessment of 84 Behavioural, Environmental and Occupational, and Metabolic Risks or Clusters of Risks, 1990–2016: A Systematic Analysis for the Global Burden of Disease Study 2016. *The Lancet* **2017**, *390* (10100), 1345–1422.
(88) MacCarty, N.; Still, D.; Ogle, D. Fuel Use and Emissions Performance of Fifty Cooking Stoves in the Laboratory and Related Benchmarks of Performance. *Energy for Sustainable Development* **2010**, *14* (3), 161–171. https://doi.org/10.1016/j.esd.2010.06.002.
(89) Diehl, J. C.; van Sprang, S.; Alexander, J.; Kersten, W. A Scalable Clean Cooking Stove Matching the Cooking Habits of Ghana and Uganda. In *2018 IEEE Global Humanitarian Technology Conference (GHTC)*; 2018; pp 1–8. https://doi.org/10.1109/GHTC.2018.8601916.
(90) Still, D.; Bentson, S.; Li, H. Results of Laboratory Testing of 15 Cookstove Designs in Accordance with the ISO/IWA Tiers of Performance. *EcoHealth* **2015**, *12* (1), 12–24. https://doi.org/10.1007/s10393-014-0955-6.
(91) Quansah, R.; Ochieng, C. A.; Semple, S.; Juvekar, S.; Emina, J.; Armah, F. A.; Luginaah, I. Effectiveness of Interventions to Reduce Indoor Air Pollution and/or Improve Health in Homes Using Solid Fuel in Lower and Middle Income Countries: Protocol for a Systematic Review. *Systematic Reviews* **2015**, *4* (1), 22. https://doi.org/10.1186/s13643-015-0012-8.
(92) *Featured Review: Assessing the effects of interventions on air pollution and health*. https://www.cochrane.org/news/featured-review-assessing-effects-interventions-air-pollution-and-health (accessed 2024-04-23).
(93) Verma, V.; Fang, T.; Xu, L.; Peltier, R. E.; Russell, A. G.; Ng, N. L.; Weber, R. J. Organic Aerosols Associated with the Generation of Reactive Oxygen Species (ROS) by Water-Soluble PM2.5. *Environ. Sci. Technol.* **2015**, *49* (7), 4646–4656. https://doi.org/10.1021/es505577w.
(94) Shafer, M. M.; Perkins, D. A.; Antkiewicz, D. S.; Stone, E. A.; Quraishi, T. A.; Schauer, J. J. Reactive Oxygen Species Activity and Chemical Speciation of Size-Fractionated

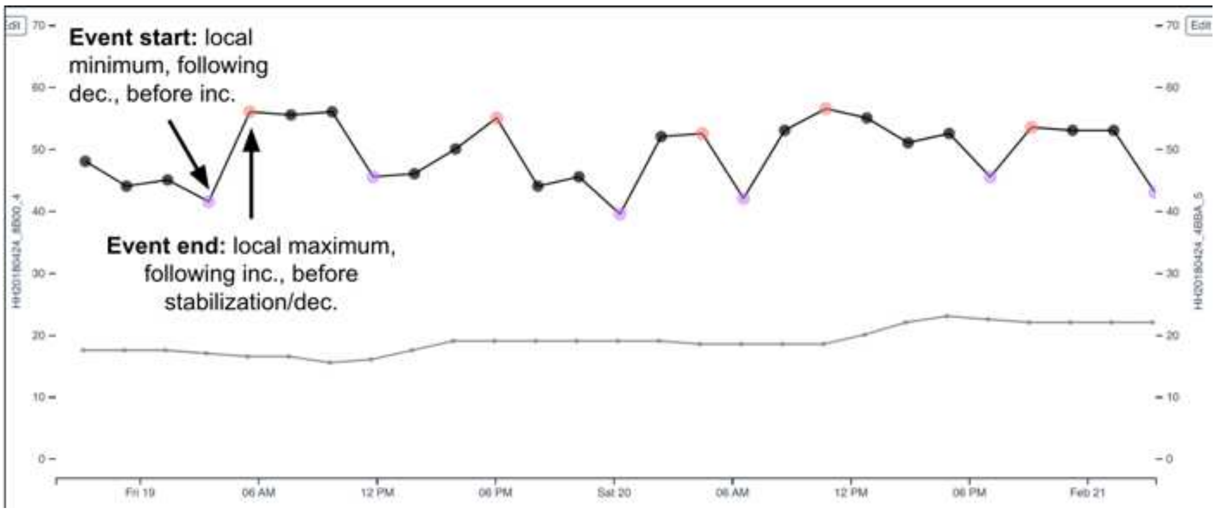
Atmospheric Particulate Matter from Lahore, Pakistan: An Important Role for Transition Metals. *J. Environ. Monit.* **2010**, *12* (3), 704–715. https://doi.org/10.1039/B915008K.
(95) Brehmer, C.; Lai, A. M.; Clark, S.; Shan, M.; Ni, K.; Ezzati, M.; Yang, X.; Baumgartner, J.; Schauer, J. J.; Carter, E. M. The Oxidative Potential of Personal and Household PM2.5 in a Rural Setting in Southwestern China. *Environ. Sci. Technol.* **2019**. https://doi.org/10.1021/acs.est.8b05120.
(96) Li, X.; Clark, S.; Floess, E.; Baumgartner, J.; Bond, T.; Carter, E. Personal Exposure to PM2.5 of Indoor and Outdoor Origin in Two Neighboring Chinese Communities with Contrasting Household Fuel Use Patterns. *Science of The Total Environment* **2021**, *800*, 149421. https://doi.org/10.1016/j.scitotenv.2021.149421.
(97) Zhao, K.; Zhang, Y.; Shang, J.; Schauer, J. J.; Huang, W.; Tian, J.; Yang, S.; Fang, D.; Zhang, D. Impact of Beijing's "Coal to Electricity" Program on Ambient PM2.5 and the Associated Reactive Oxygen Species (ROS). *Journal of Environmental Sciences* **2023**, *133*, 93–106. https://doi.org/10.1016/j.jes.2022.06.038.
(98) Lai, A.; Lee, M.; Carter, E.; Chan, Q.; Elliott, P.; Ezzati, M.; Kelly, F.; Yan, L.; Wu, Y.; Yang, X.; Zhao, L.; Baumgartner, J.; Schauer, J. J. Chemical Investigation of Household Solid Fuel Use and Outdoor Air Pollution Contributions to Personal PM2.5 Exposures. *Environ. Sci. Technol.* **2021**, *55* (23), 15969–15979. https://doi.org/10.1021/acs.est.1c01368.
(99) *Residential Coal Combustion as a Source of Levoglucosan in China | Environmental Science & Technology*. https://pubs.acs.org/doi/full/10.1021/acs.est.7b05858?casa_token=5Iio34twuzsAAAAA%3AbDkbS6d7IYSqEwTHQtrLSxunoEKnn9LVE8O1NpEHzPFv4IFCCxRKw7J-A43GSK5O5qZ-rE9XXd-dIhs (accessed 2024-11-07).
(100) Hopke, P. K.; Dai, Q.; Li, L.; Feng, Y. Global Review of Recent Source Apportionments for Airborne Particulate Matter. *Science of The Total Environment* **2020**, *740*, 140091. https://doi.org/10.1016/j.scitotenv.2020.140091.
(101) Zhu, Y.; Huang, L.; Li, J.; Ying, Q.; Zhang, H.; Liu, X.; Liao, H.; Li, N.; Liu, Z.; Mao, Y.; Fang, H.; Hu, J. Sources of Particulate Matter in China: Insights from Source Apportionment Studies Published in 1987–2017. *Environment International* **2018**, *115*, 343–357. https://doi.org/10.1016/j.envint.2018.03.037.
(102) Athey, S.; Wager, S. Estimating Treatment Effects with Causal Forests: An Application. *Observational studies* **2019**, *5* (2), 37–51.
(103) Hahn, P. R.; Murray, J. S.; Carvalho, C. M. Bayesian Regression Tree Models for Causal Inference: Regularization, Confounding, and Heterogeneous Effects (with Discussion). *Bayesian Analysis* **2020**, *15* (3), 965–1056.
(104) Volckens, J.; Quinn, C.; Leith, D.; Mehaffy, J.; Henry, C. S.; Miller-Lionberg, D. Development and Evaluation of an Ultrasonic Personal Aerosol Sampler. *Indoor Air* **2017**, *27* (2), 409–416. https://doi.org/10.1111/ina.12318.
(105) L'Orange, C.; Neymark, G.; Carter, E.; Volckens, J. A High-Throughput, Robotic System for Analysis of Aerosol Sampling Filters. *Aerosol Air Qual. Res.* **2021**, *21* (11), 210037. https://doi.org/10.4209/aaqr.210037.
(106) Li, Y.; Schwandner, F. M.; Sewell, H. J.; Zivkovich, A.; Tigges, M.; Raja, S.; Holcomb, S.; Molenar, J. V.; Sherman, L.; Archuleta, C. Observations of Ammonia, Nitric Acid, and Fine Particles in a Rural Gas Production Region. *Atmospheric Environment* **2014**, *83*, 80–89.

(107) Ahmed, T.; Dutkiewicz, V. A.; Shareef, A.; Tuncel, G.; Tuncel, S.; Husain, L. Measurement of Black Carbon (BC) by an Optical Method and a Thermal-Optical Method: Intercomparison for Four Sites. *Atmospheric Environment* **2009**, *43* (40), 6305–6311. https://doi.org/10.1016/j.atmosenv.2009.09.031.
(108) Khuzestani, R. B.; Schauer, J. J.; Wei, Y.; Zhang, Y.; Zhang, Y. 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* **2017**, *149*, 84–94. https://doi.org/10.1016/j.atmosenv.2016.11.002.
(109) Norris, G. A.; Duvall, R.; Brown, S. G.; Bai, S. EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide Prepared for the US Environmental Protection Agency Office of Research and Development, Washington, DC. *Inc., Petaluma* **2014**.
(110) Brown, S. G.; Eberly, S.; Paatero, P.; Norris, G. A. Methods for Estimating Uncertainty in PMF Solutions: Examples with Ambient Air and Water Quality Data and Guidance on Reporting PMF Results. *Science of The Total Environment* **2015**, *518–519*, 626–635. https://doi.org/10.1016/j.scitotenv.2015.01.022.
(111)Dai, Q.; Liu, B.; Bi, X.; Wu, J.; Liang, D.; Zhang, Y.; Feng, Y.; Hopke, P. K. Dispersion Normalized PMF Provides Insights into the Significant Changes in Source Contributions to PM2. 5 after the COVID-19 Outbreak. *Environmental Science & Technology* **2020**, *54* (16), 9917–9927.
(112) Bonander, C.; Svensson, M. Using Causal Forests to Assess Heterogeneity in Cost-Effectiveness Analysis. *Health Economics* **2021**, *30* (8), 1818–1832. https://doi.org/10.1002/hec.4263.
(113) Miller, S. Causal Forest Estimation of Heterogeneous and Time-Varying Environmental Policy Effects. *Journal of Environmental Economics and Management* **2020**, *103*, 102337. https://doi.org/10.1016/j.jeem.2020.102337.
(114) Rosenthal, J.; Quinn, A.; Grieshop, A. P.; Pillarisetti, A.; Glass, R. I. Clean Cooking and the SDGs: Integrated Analytical Approaches to Guide Energy Interventions for Health and Environment Goals. *Energy for Sustainable Development* **2018**, *42*, 152–159.
(115) Ezzati, M.; Baumgartner, J. C. Household Energy and Health: Where next for Research and Practice? *The Lancet* **2017**, *389* (10065), 130–132. https://doi.org/10.1016/S0140-6736(16)32506-5.
(116) Lai, P. S.; Lam, N. L.; Gallery, B.; Lee, A. G.; Adair-Rohani, H.; Alexander, D.; Balakrishnan, K.; Bisaga, I.; Chafe, Z. A.; Clasen, T.; Díaz-Artiga, A.; Grieshop, A.; Harrison, K.; Hartinger, S. M.; Jack, D.; Kaali, S.; Lydston, M.; Mortimer, K. M.; Nicolaou, L.; Obonyo, E.; Okello, G.; Olopade, C.; Pillarisetti, A.; Pinto, A. N.; Rosenthal, J. P.; Schluger, N.; Shi, X.; Thompson, C.; Thompson, L. M.; Volckens, J.; Williams, K. N.; Balmes, J.; Checkley, W.; Ozoh, O. B. Household Air Pollution Interventions to Improve Health in Low- and Middle-Income Countries: An Official American Thoracic Society Research Statement. *Am J Respir Crit Care Med* **2024**, *209* (8), 909–927. https://doi.org/10.1164/rccm.202402-0398ST.
(117) Shang, J.; Khuzestani, R. B.; Tian, J.; Schauer, J. J.; Hua, J.; Zhang, Y.; Cai, T.; Fang, D.; An, J.; Zhang, Y. Chemical Characterization and Source Apportionment of PM2.5 Personal Exposure of Two Cohorts Living in Urban and Suburban Beijing. *Environmental Pollution* **2019**, *246*, 225–236. https://doi.org/10.1016/j.envpol.2018.11.076.
(118) Zong, Z.; Wang, X.; Tian, C.; Chen, Y.; Fu, S.; Qu, L.; Ji, L.; Li, J.; Zhang, G. PMF and PSCF Based Source Apportionment of PM2.5 at a Regional Background Site in North

China. *Atmospheric Research* **2018**, *203*, 207–215. https://doi.org/10.1016/j.atmosres.2017.12.013.
(119) Li, H.; Wang, Q.; Yang, M.; Li, F.; Wang, J.; Sun, Y.; Wang, C.; Wu, H.; Qian, X. Chemical Characterization and Source Apportionment of PM2.5 Aerosols in a Megacity of Southeast China. *Atmospheric Research* **2016**, *181*, 288–299. https://doi.org/10.1016/j.atmosres.2016.07.005.
(120) Wang, Y.; Zhuang, G.; Tang, A.; Yuan, H.; Sun, Y.; Chen, S.; Zheng, A. The Ion Chemistry and the Source of PM2.5 Aerosol in Beijing. *Atmospheric Environment* **2005**, *39* (21), 3771–3784. https://doi.org/10.1016/j.atmosenv.2005.03.013.
(121) Zhang, R.; Jing, J.; Tao, J.; Hsu, S.-C.; Wang, G.; Cao, J.; Lee, C. S. L.; Zhu, L.; Chen, Z.; Zhao, Y.; Shen, Z. Chemical Characterization and Source Apportionment of PM_{2.5} in Beijing: Seasonal Perspective. *Atmospheric Chemistry and Physics* **2013**, *13* (14), 7053–7074. https://doi.org/10.5194/acp-13-7053-2013.
(122) Lei, W.; Zhang, L.; Xu, J.; Liu, Z.; Xin, J.; Li, X.; Zhao, W. Spatiotemporal Variations and Source Apportionment of Metals in Atmospheric Particulate Matter in Beijing and Its Surrounding Areas. *Atmospheric Pollution Research* **2021**, *12* (11), 101213. https://doi.org/10.1016/j.apr.2021.101213.
(123) Yu, S.; Liu, W.; Xu, Y.; Yi, K.; Zhou, M.; Tao, S.; Liu, W. Characteristics and Oxidative Potential of Atmospheric PM2.5 in Beijing: Source Apportionment and Seasonal Variation. *Science of The Total Environment* **2019**, *650*, 277–287. https://doi.org/10.1016/j.scitotenv.2018.09.021.
(124) Gao, J.; Wang, K.; Wang, Y.; Liu, S.; Zhu, C.; Hao, J.; Liu, H.; Hua, S.; Tian, H. Temporal-Spatial Characteristics and Source Apportionment of PM2.5 as Well as Its Associated Chemical Species in the Beijing-Tianjin-Hebei Region of China. *Environmental Pollution* **2018**, *233*, 714–724. https://doi.org/10.1016/j.envpol.2017.10.123.
(125) Piedrahita, R.; Kanyomse, E.; Coffey, E.; Xie, M.; Hagar, Y.; Alirigia, R.; Agyei, F.; Wiedinmyer, C.; Dickinson, K. L.; Oduro, A.; Hannigan, M. Exposures to and Origins of Carbonaceous PM2.5 in a Cookstove Intervention in Northern Ghana. *Science of The Total Environment* **2017**, *576*, 178–192. https://doi.org/10.1016/j.scitotenv.2016.10.069.
(126) Lai, A. M.; Carter, E.; Shan, M.; Ni, K.; Clark, S.; Ezzati, M.; Wiedinmyer, C.; Yang, X.; Baumgartner, J.; Schauer, J. J. Chemical Composition and Source Apportionment of Ambient, Household, and Personal Exposures to PM2.5 in Communities Using Biomass Stoves in Rural China. *Science of The Total Environment* **2019**, *646*, 309–319. https://doi.org/10.1016/j.scitotenv.2018.07.322.
(127) Mollinedo, E. Source Apportionment of Fine Particulate Matter (PM2.5) Personal Exposure Samples from the Household Air Pollution Intervention Network (HAPIN) Trial in Jalapa, Guatemala, 2024. https://virtual.oxfordabstracts.com/event/20221/submission/392.

APPENDIX A

Appendix A1. Example of a manually labeled biomass kang surface temperature stove use event. The y-axis is the measured temperature, the x-axis is time of day. The darker black line is the stove temperature, and the lighter grey line is the indoor temperature.



Appendix A2. Short-term kang stove use event methods and results

Appendix A2.1. Short-term kang event usage modeling methods

For the short-term kang models, 15% of files were selected to hand label event starts and ends. Prior to hand-labeling the data, we took 16-minute averages and decreased the resolution of the data from 2-minutes to 16-minutes. We found that 2-minute data was noisy, and a 16-minute average visually kept the same general variability in the data while reducing some of the noise. Due to the relatively small number of events in 15% of the short-term files (n = 64), the XGBoost models to predict the probability that any given observation was a start or an end was tuned and trained using 2-fold 1000-repeat cross validation. The variables included in the model, model evaluation, and post-prediction cleaning were identical to that of the long-term kang stoves.

Appendix A2.2. Short-term kang event usage modeling results

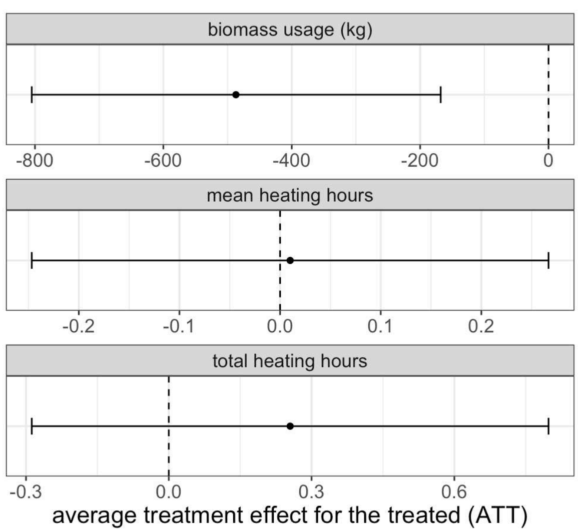
The sensitivity, specificity, and F-score for the short-term kang XGBoost models predicting the probability than an observation is an event start or an event end are shown in the table below. A cutoff of 0.60 for the start model and 0.25 for the end model were determined by taking the cutoff value where the F-score was the largest value. The start model performed better than the end model, with F-scores of 0.996 and 0.665 respectively. The start model was better at identifying true positive events (sensitivity: 99.72%) compared to the end model (sensitivity: 66.48%), while the end model was better at predicting true negative events (specificity: 99.49%) than the start model (specificity: 68.16%).

	Short-term[a] kangs	
	Start	End
Sensitivity	99.72%	66.48%
Specificity	68.16%	99.49%
F-score	99.62%	66.51%

[a]Data collected at 2-minute intervals averaged and reduced to 16-minutes from Nov-March

While we did apply a cross-validation approach to training the model to utilize all the data, the short-term models still generally performed worse than the long-term models. One reason for the poorer short-term model performance is the limited number of events that the model was trained on (n = 64). The relative number of events to non-events (~3% of the short-term data was starts or ends) also likely played a role in poorer model performance because classification models can perform poorly with highly imbalanced data. However, the cross-validation method for model training should have mitigated some of the impacts of data imbalance on model performance. We suggest that future studies that apply similar methods as this work to predict event starts and ends in higher resolution data hand-label a higher percentage of data to ensure the model has enough data to be trained on.

Appendix A3. Difference-in-difference analysis for self-reported biomass usage, mean biomass heating hours, and total biomass heating hours.



The impacts of the CHP policy on biomass usage, mean heating hours, and total heating hours using the same extended two-way fixed effects model outlined in the main text are shown in Figure S2. The CHP policy reduced total self-reported biomass usage by 487 kg (95% CI: 805-168 kg) and was significant at an alpha of 0.05. This indicates that participants generally used less biomass after the policy was implemented. The most likely reason why this could have occurred is that the new heat pump was better, or sufficient, for household heating so less biomass was needed to heat the home. The mean heating hours and total heating hours for biomass usage did not change because of the CHP. The heating hour models were adjusted for the price of coal, income, age, sex, marital status, education level, household size, number of children under 5, household area, and agricultural assets. Taken together with the reduction in biomass usage due to the policy, these results suggest that participants were still using their wood burning devices the same amount but were using less biomass when they were using them.

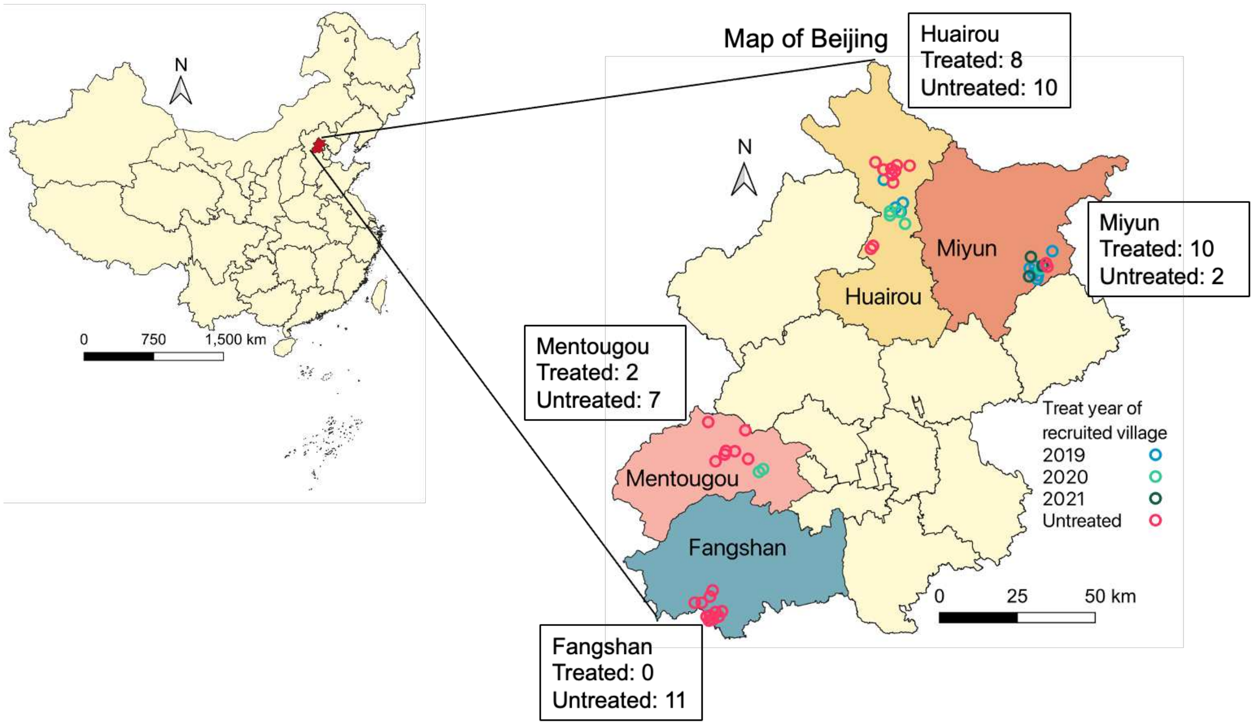
APPENDIX B

Appendix B1. Details about participant recruitment.

In Wave-1, we recruited 494 participants for personal exposure measurements, and 72% of them underwent measurements in Wave-2. Additionally, 65% and 82% of participants who underwent personal exposure measurements in Wave-1 and Wave-2 also participated in Wave-4. Sensor-reported indoor $PM_{2.5}$ measurements were conducted in 300 households in both Wave-2 and Wave-4, and 152 and 151 indoor $PM_{2.5}$ filter samples were collected in Wave-2 and Wave-4, respectively.

	Overall			Personal exposure			Indoor low-cost sensor			Indoor filter	
	Wave-1	Wave-2	Wave-4	Wave-1	Wave-2	Wave-4	Wave-1	Wave-2	Wave-4	Wave-2	Wave-4
New recruitment	977	189	68	494	135	68	0	300	52	152	30
Households from Wave-1	/	866	782	/	363	322	/	0	0	/	/
Households from Wave-2	/	/	162	/	/	109	/	/	248	/	121
Total recruitment	977	1055	1012	494	498	499	0	300	300	152	151

Appendix B2. Map illustrating the geographical distribution of the recruited villages in this study within Beijing Municipality. Each circle represents one recruited village, with colors indicating the year of involvement in the Coal-to-Clean-Energy Policy (CCEP). The CCEP treatment status of villages in each district in the final field season is listed in the black frame. The map plots were created using QGIS3.30.



Appendix B3. Covariate indices developed from household questionnaire

The three primary heating fuel types in our recruited households were electricity, coal, and wood. Most of them did not only have a primary fuel but also had another type of fuels as a supplementary. Based on the used heating fuel types, we classified the heating energy use patterns of our recruited households into eight categories, i.e., electricity exclusively, coal + electricity, wood + electricity, wood + coal + electricity, coal exclusively, wood exclusively, wood + coal, and no heating.

Cigarette smoking is an important indoor air pollution emission source besides solid fuel combustion. To gather data on this issue, we inquired whether the recruited participants smoked and if not, whether they lived with someone who smoked. This information was then merged into a variable indicating whether there was a smoker in the household.

We also collected information on household income, pension, assets, house area, and owned agricultural and forest land area, which could indicate the household socioeconomical status. We applied these household assets information to principle component analysis (PCA) to create a composite (wealth) index to indicate the household wealth index [70].

Appendix B4. Indoor $PM_{2.5}$ sensor exclusion criteria

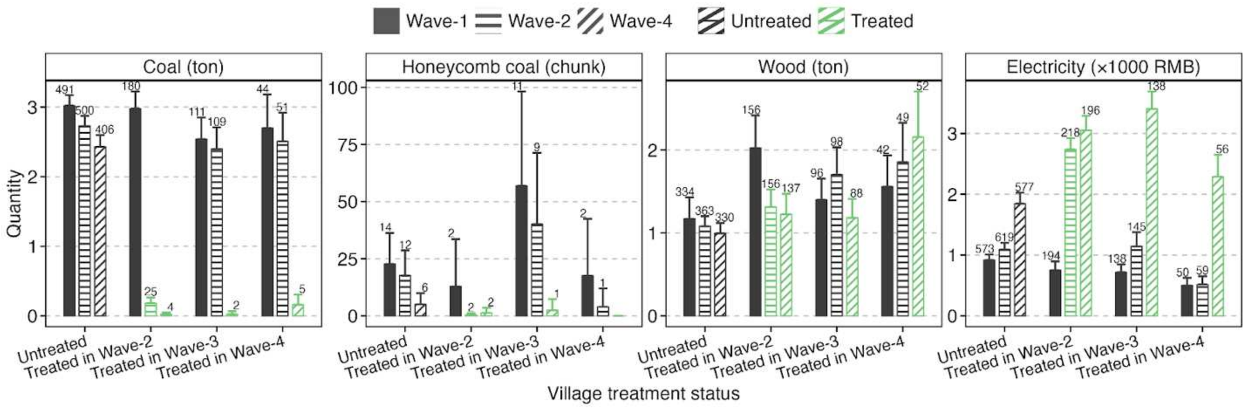
Some PM2.5 sensors encountered issues during their extended deployment in the field (4–6 months), leading to incomplete data collection. Three main problems were identified: (1) power failures, affecting 35 sensors in S2 and 43 in S4; (2) sensor malfunctions, with 16 and 15 broken sensors in S2 and S4, respectively; and (3) loss of sensors, occurring to 2 in S2 and 11 in S4.

Appendix B5. $PM_{2.5}$ sensor calibration methods

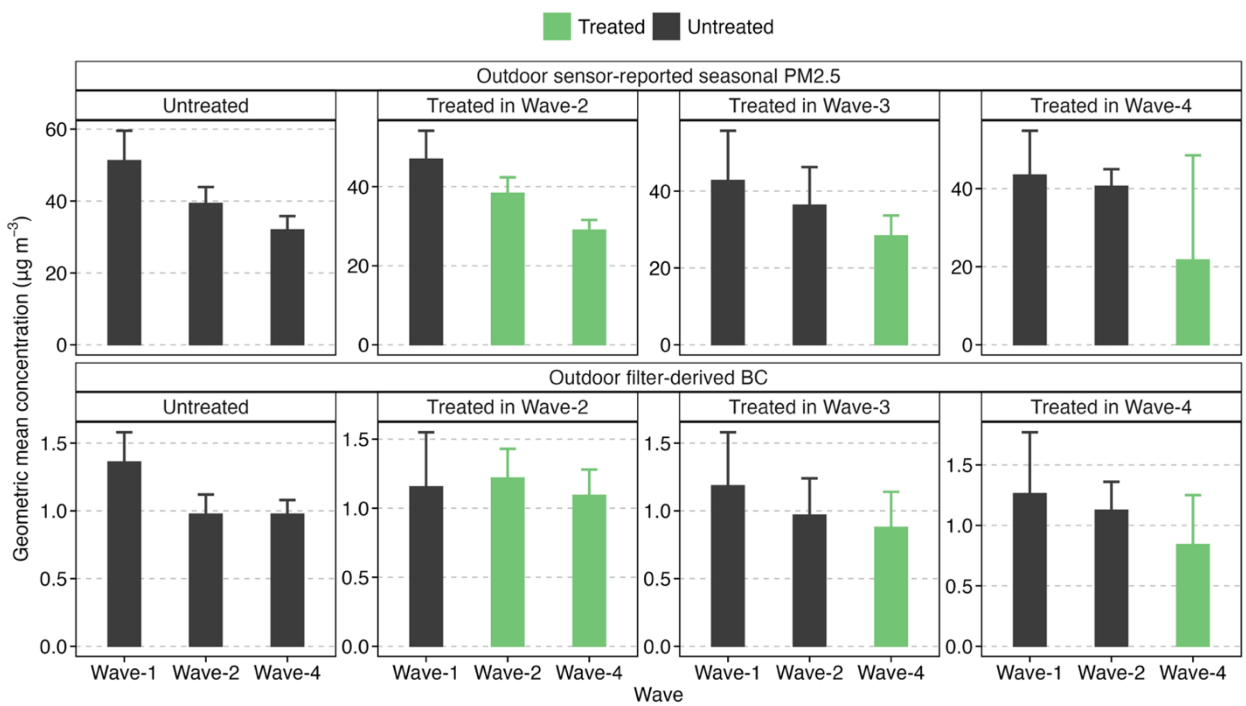
Before and after each field campaign, we deployed all PM sensors at a rooftop of Peking University campus (PKU, urban site) and/or the University of Chinese Academy of Sciences campus (UCAS, peri-urban site) for 7 to 10 days. This was done to compare our sensor data with the hourly Federal Equivalent Methods (FEM) observations by the Thermo Electron Synchronized Hybrid Ambient Real-Time Particulate Monitor (model 5030) at Peking University campus or Tapered Element Oscillating Microbalance Method (TEOM) from the closest China National Environmental Monitoring Centre (CNEMC) (publicly available at <https://quotsoft.net/air/>). The distances from the PKU and UCAS campuses to the Environmental Monitoring Station are 1.7 and 9.9 km, respectively. Hourly sensor-reported $PM_{2.5}$ concentrations were strongly correlated with those measured by the FEM instruments, with all Spearman correlation coefficients (*rho*) exceeding 0.8 (Li et al., in preparation).

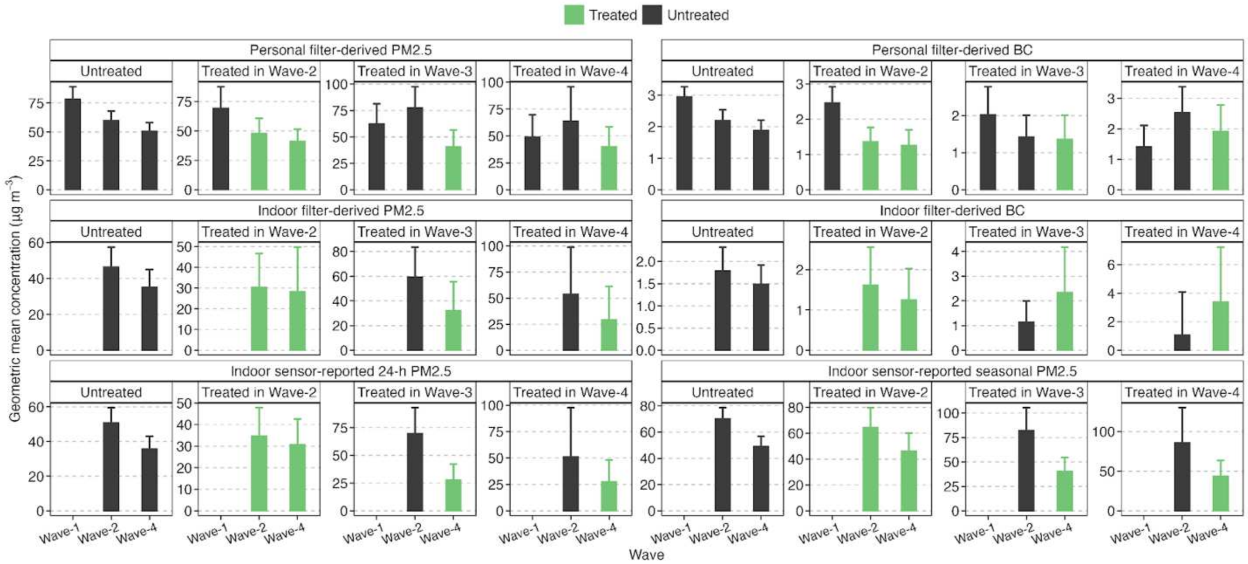
To adjust the sensor-reported outdoor and indoor $PM_{2.5}$ data, we established linear regression models between the filter-derived $PM_{2.5}$ mass concentrations (i.e., the reference concentrations) and the sensor-reported $PM_{2.5}$ concentrations averaged over the same sampling period as the filter samples. The slopes of the models were used as the adjustment factors for the sensor-reported $PM_{2.5}$ concentrations (Li et al., in preparation). Separate calibration models were built for outdoor and indoor sensors by wave.

Appendix B6. Reported heating fuel consumption. Coal consumption is estimated for winter use in tons; honeycomb coal consumption, in chunks; wood consumption, for annual use in tons; and electricity use, for winter use based on monetary payments. Bar colors indicate the treatment status of villages in each study wave: black for untreated and green for treated. Bar filling patterns represent the study wave: solid for Wave-1, horizontal stripes for Wave-2, and diagonal stripes for Wave-4. The height of each bar reflects the mean fuel consumption, with error bars showing the upper bound of the 95%CI. Numbers displayed above the bars indicate the number of households reporting usage of the corresponding fuel.

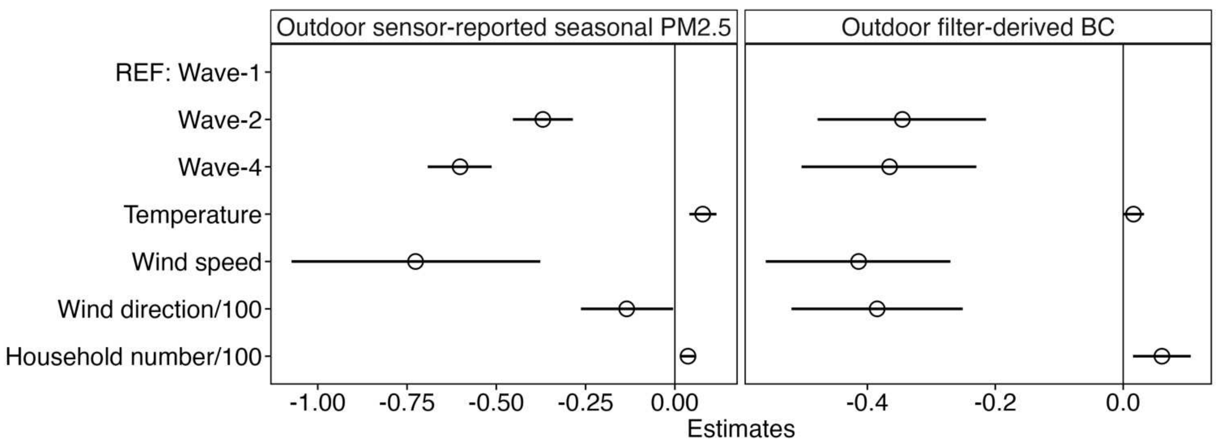


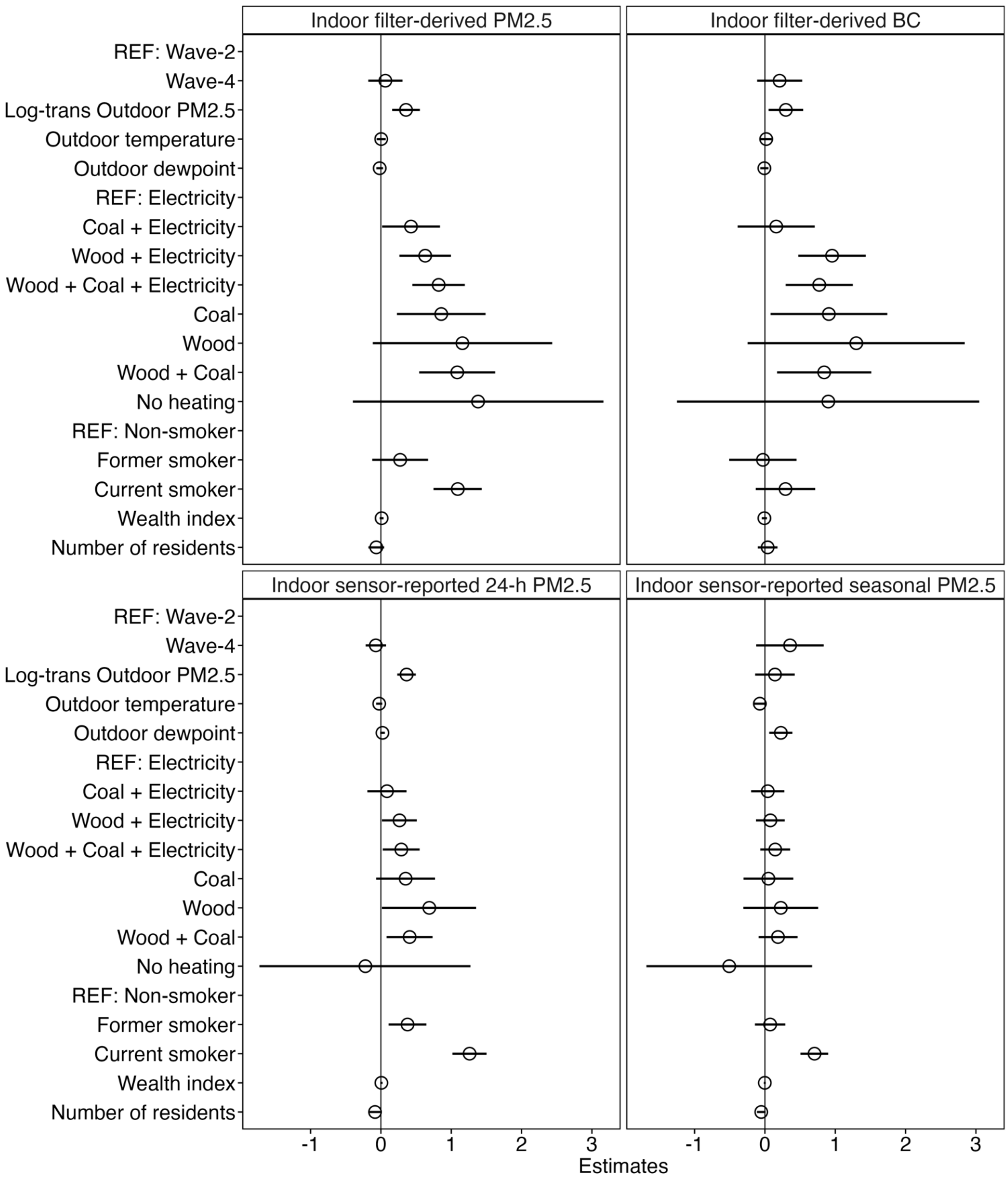
Appendix B7. Geometric mean and 95% confidence intervals for outdoor seasonal $PM_{2.5}$, outdoor black carbon (BC), personal $PM_{2.5}$ exposure, personal black carbon exposure, indoor seasonal $PM_{2.5}$, and indoor black carbon.



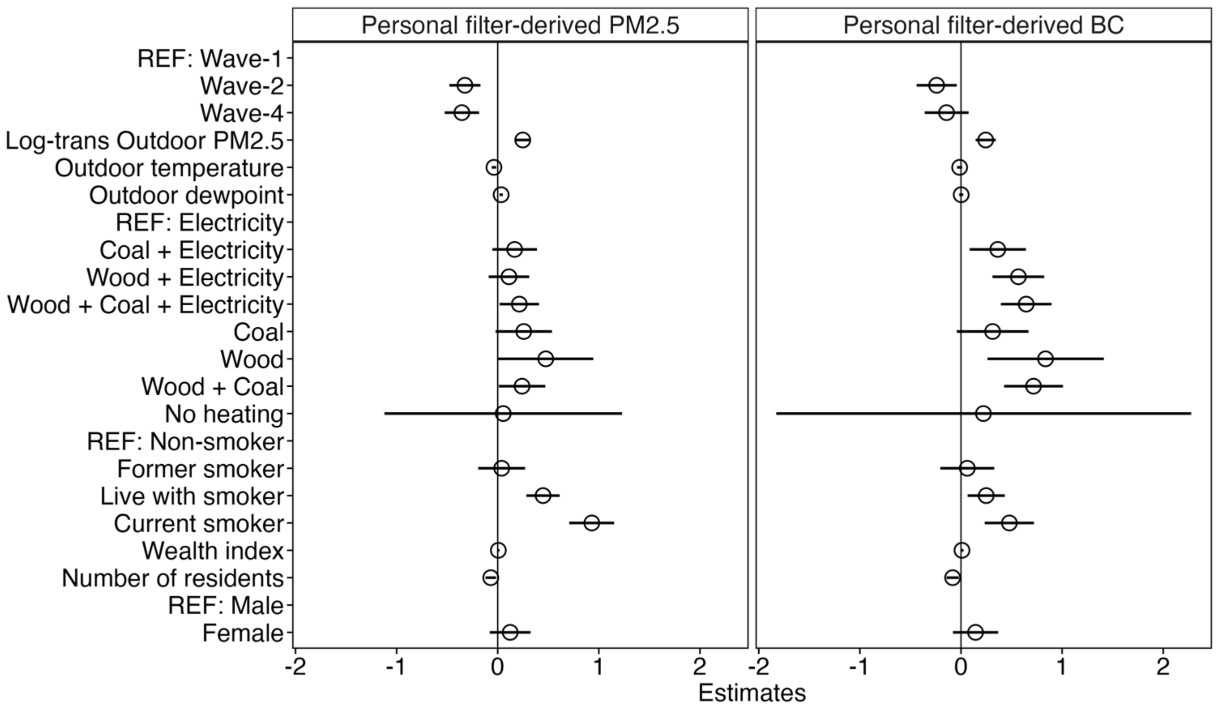


Appendix B8. Associations between selected variables and outdoor seasonal $PM_{2.5}$, outdoor black carbon (BC), personal $PM_{2.5}$ exposure, personal black carbon exposure, indoor seasonal $PM_{2.5}$, and indoor black carbon.

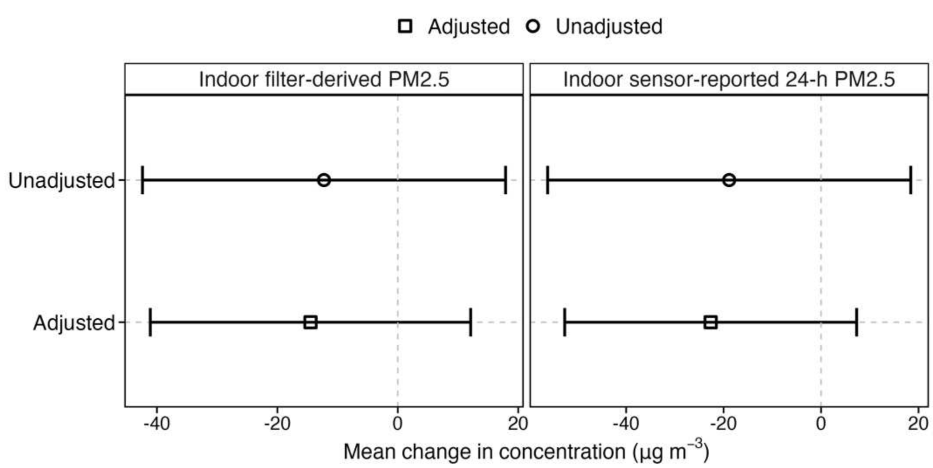




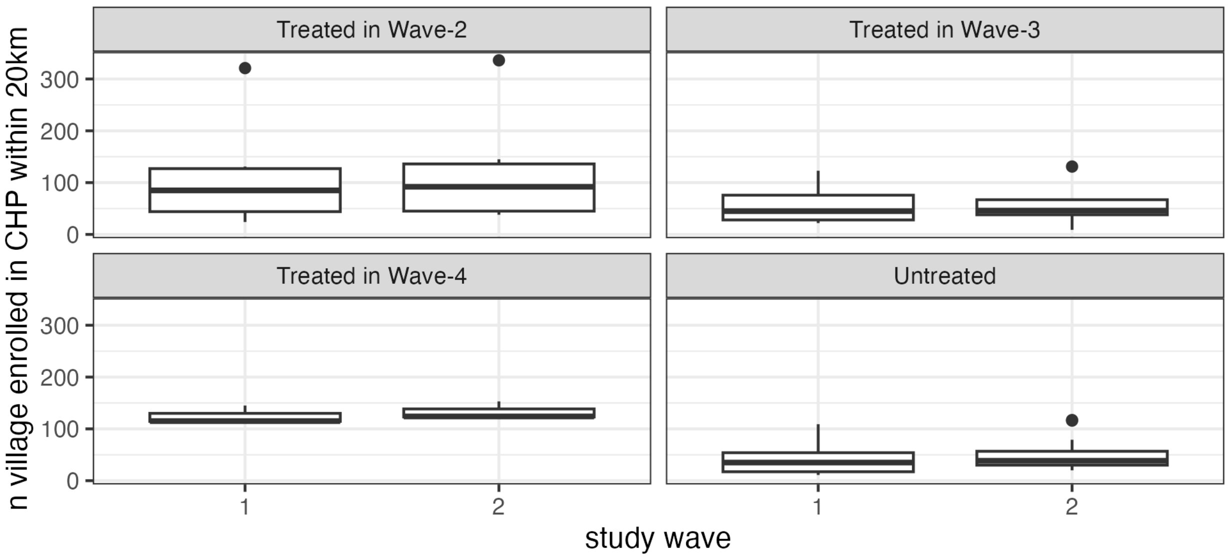
Indoor filter-derived PM2.5
Indoor filter-derived BC
Indoor sensor-reported 24-h PM2.5
Indoor sensor-reported seasonal PM2.5
REF: Wave-2
Wave-4
Log-trans Outdoor PM2.5
Outdoor temperature
Outdoor dewpoint
REF: Electricity
Coal + Electricity
Wood + Electricity
Wood + Coal + Electricity
Coal
Wood
Wood + Coal
No heating
REF: Non-smoker
Former smoker
Current smoker
Wealth index
Number of residents
-1
0
1
2
3
Estimates



Appendix B9. The effect of the CCEP on 24-h indoor $PM_{2.5}$ using extended two-way fixed effect difference-in-difference analysis (DID). The "*Unadjusted"* y-axis label refers to the ETWFE model without any control variables, while the "*Adjusted"* y-axis label refers to the ETWFE model incorporating control variables. Dots represent the mean treatment effect, and error bars denote the 95% CIs.



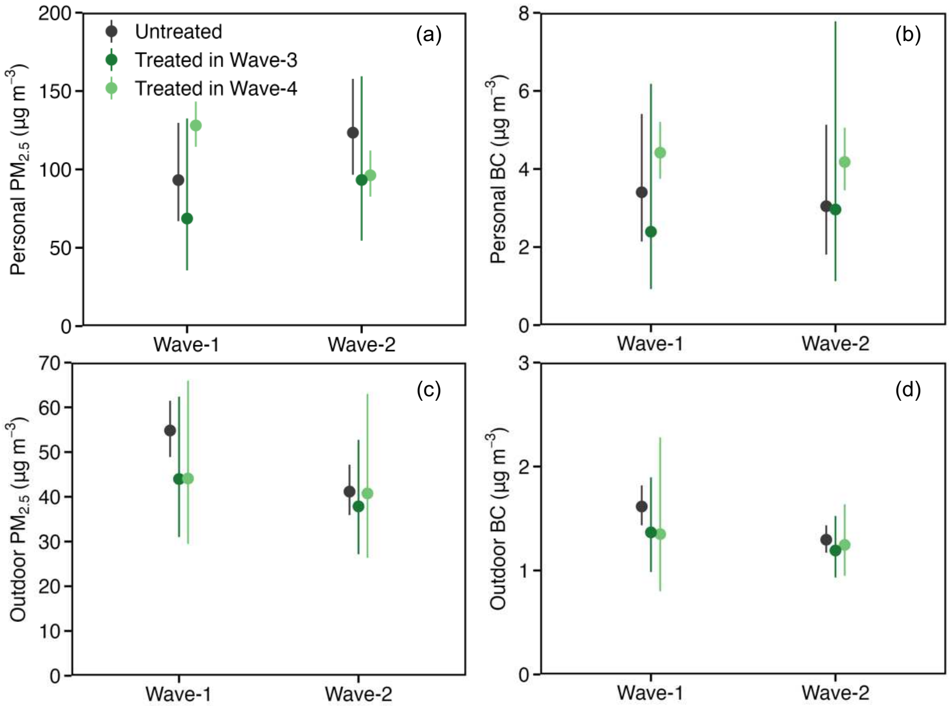
Appendix B10. Number of villages within 20km of our study villages that were enrolled in the CHP.



Appendix B11. Impacts of including a district fixed-effect on the treatment effect estimation for personal $PM_{2.5}$ and BC exposure, 24-hr indoor $PM_{2.5}$, and seasonal average indoor $PM_{2.5}$.

	$PM_{2.5}$			BC		
Model	**Observations**	**ATT**	**95% CIs**	**Observations**	**ATT**	**95% CIs**
Outdoor	Seasonal mean $PM_{2.5}$			Filter-derived seasonal BC		
Adjusted model	139	0.4	-3.1, 4.0	699	0.2	-0.1, 0.5
Excluding Fangshan villages	107	-1.1	-5.7, 3.4	540	0.3	-0.2, 0.7
With Wave-3 data	189	0.4	-1.5, 2.4	/	/	/
Indoor	Seasonal mean $PM_{2.5}$			Filter-derived 24-h BC		
Adjusted model	366	-22.2	-40.3, -4.2	189	2.9	0.7, 5.0
With district fixed effect	366	-22.6	-40.1, -5.1	189	3.0	0.8, 5.2
Unadjusted	366	-30.9	-53.2, -8.7	189	2.6	0.4, 4.7
With Wave-3 data	523	-35.4	-57.2, -13.7	/	/	/
Personal exposure	Filter-derived 24-h $PM_{2.5}$			Filter-derived 24-h BC		
Adjusted model	1270	-3.6	-23.6, 16.5	1161	-0.5	-1.5, 0.5
No wave fixed-effect	1270	-27.2	-48.3, -6.1	1161	-0.6	-1.5, 0.2
No cohort fixed-effect	1270	-9.3	-28.6, 10.0	1161	-1.4	-2.6, -0.3
No fixed effect	1270	-24.1	-41.5, -6.7	1161	-1.4	-2.2, -0.5
With district fixed-effect	1270	-3.8	-23.0, 15.5	1161	-0.4	-1.4, 0.7

Appendix B12. Pre-intervention air pollution trends for outdoor and personal $PM_{2.5}$ and black carbon (BC) exposure.



APPENDIX C

APPENDIX C1. Arithmetic mean concentration (100 x ug m^{-3}) and 95% confidence intervals for measured chemical species outdoor and personal $PM_{2.5}$ exposure measurements included in source apportionment.

Outdoor

wave	Al	Fe	Pb	Si	Chloride	EC	ws-Na	Ammonium	Nitrate
1	110 (108-113)	85.2 (82.9-87.5)	6.26 (5.82-6.71)	139 (136-142)	56.1 (54.5-57.8)	123 (121-125)	11.4 (10.8-12)	136 (133-138)	225 (222-227)
2	66.7 (65.5-67.8)	57.9 (57-58.9)	4.32 (4.06-4.57)	76.7 (75.4-77.9)	32.2 (31.5-32.9)	112 (111-113)	9.28 (8.88-9.68)	152 (151-154)	259 (257-261)
4	60.9 (59.8-62.1)	64.1 (63-65.2)	7.69 (7.27-8.11)	128 (126-129)	15.6 (14.9-16.4)	103 (101-104)	9.23 (8.76-9.7)	78.3 (77.3-79.4)	146 (145-148)
wave	ns-S	OC	Sulfate	wi-Ca	wi-K	wi-Mg	ws-Ca	ws-K	ws-Mg
1	87 (85.3-88.7)	1080 (1070-1080)	223 (220-225)	79.4 (77.2-81.7)	81.4 (79.5-83.3)	39.8 (38.6-41)	33.5 (32.5-34.5)	52.4 (50.8-54)	6.98 (6.41-7.55)
2	70 (69.2-70.9)	901 (898-904)	206 (205-208)	40.7 (39.8-41.6)	33.7 (33-34.3)	21.2 (20.6-21.7)	23.5 (22.8-24.2)	25.4 (24.9-25.9)	3.19 (2.97-3.41)
4	75.9 (74.9-76.9)	649 (645-653)	132 (131-134)	79 (77.6-80.4)	37.8 (37-38.6)	38.6 (37.7-39.4)	15.1 (14.3-15.9)	17.9 (17.4-18.4)	2.96 (2.72-3.2)

Personal exposure

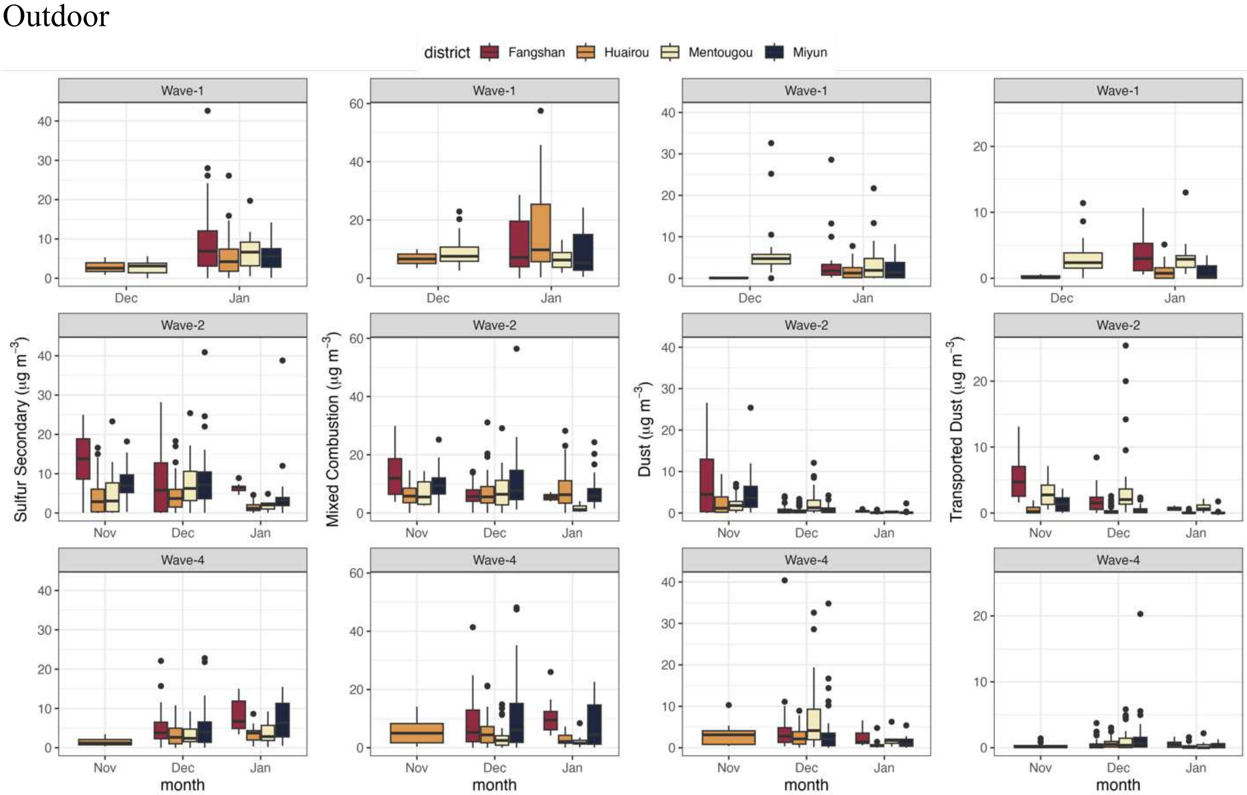
wave	Al	Fe	Pb	Si	Chloride	EC	ws-Na	Ammonium	Nitrate
1	203 (202-204)	40 (39.4-40.6)	15.8 (15.4-16.1)	77.2 (75.9-78.4)	70.2 (69-71.5)	241 (239-243)	8.75 (8.36-9.14)	60.5 (59.5-61.5)	151 (149-152)
2	213 (212-214)	57.1 (55.9-58.4)	13.7 (13.5-14)	93.5 (92-95)	43.8 (42.7-45)	219 (218-221)	12.3 (11.8-12.7)	87.8 (86.4-89.1)	235 (232-237)
4	27.4 (26.7-28.1)	33.6 (32.9-34.3)	12.8 (12.4-13.1)	52.1 (51.2-53)	34.6 (33.6-35.5)	194 (192-196)	10.8 (10.3-11.2)	31.1 (30.2-31.9)	59.1 (58.1-60.1)
wave	ns-S	OC	Sulfate	wi-Ca	wi-K	wi-Mg	ws-Ca	ws-K	ws-Mg
1	46.3 (45.4-47.2)	3740 (3730-3740)	129 (127-130)	46.8 (45.7-47.9)	88.2 (87-89.4)	25.6 (25-26.2)	57.1 (56.4-57.9)	49 (48-50)	4 (3.75-4.24)
2	44.1 (43.2-45)	3430 (3420-3440)	126 (125-128)	81.6 (79.8-83.5)	72.8 (71.6-74)	30.4 (29.7-31.1)	42.8 (41.9-43.8)	43.8 (42.9-44.7)	3.77 (3.49-4.04)
4	27.5 (26.7-28.3)	2680 (2680-2690)	77.6 (76.5-78.7)	48.4 (47.6-49.2)	67.7 (66.5-68.9)	26.5 (26-27)	15.7 (15.1-16.2)	28.7 (27.9-29.5)	2.06 (1.81-2.3)

Appendix C2. Positive Matrix Factorization source apportionment model diagnostics.

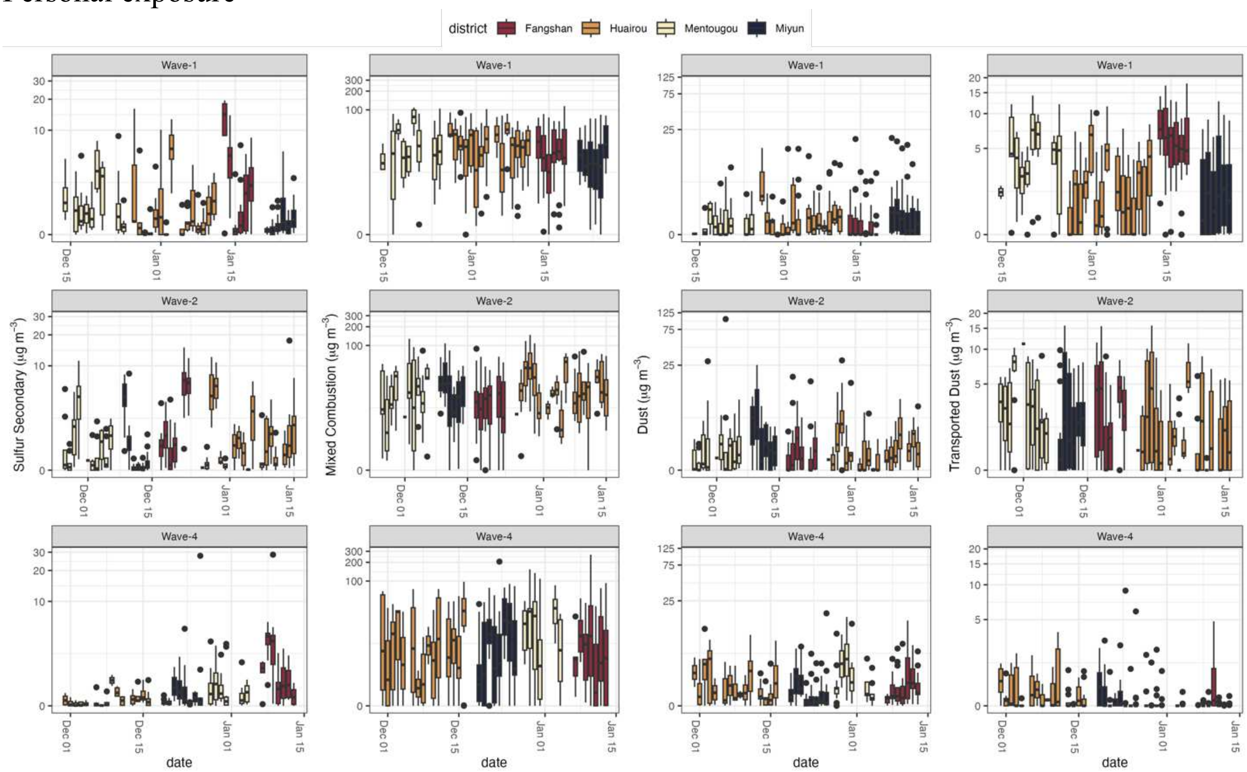
Diagnostic	3	4	5	6
Q_{exp}	27936	26052	24168	22284
Q_{true}	187681	147796	123236	100316
Q_{robust}	174407	139910	117082	95932.5
Q_r/Q_{exp}	6.24	5.37	4.84	4.30
$Q/Q_{exp} > 6$	wi-Ca, ns-S, ws-Na, ws-Ca, Al, Cl, Pb	ns-S, Na, Al, Cl, Pb, Nitrate	Nitrate, ws-Na, Al, Chloride	Nitrate, ws-Na, Al
DISP % dQ	<0.1%	<0.1%	<0.1%	<0.1%
DISP swaps	0	0	0	0
BS mapping < 100%	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%

Appendix C3. Outdoor and personal exposure to the source contributions (μg m^{-3}) aggregated by day of year, study wave, and colored by district.

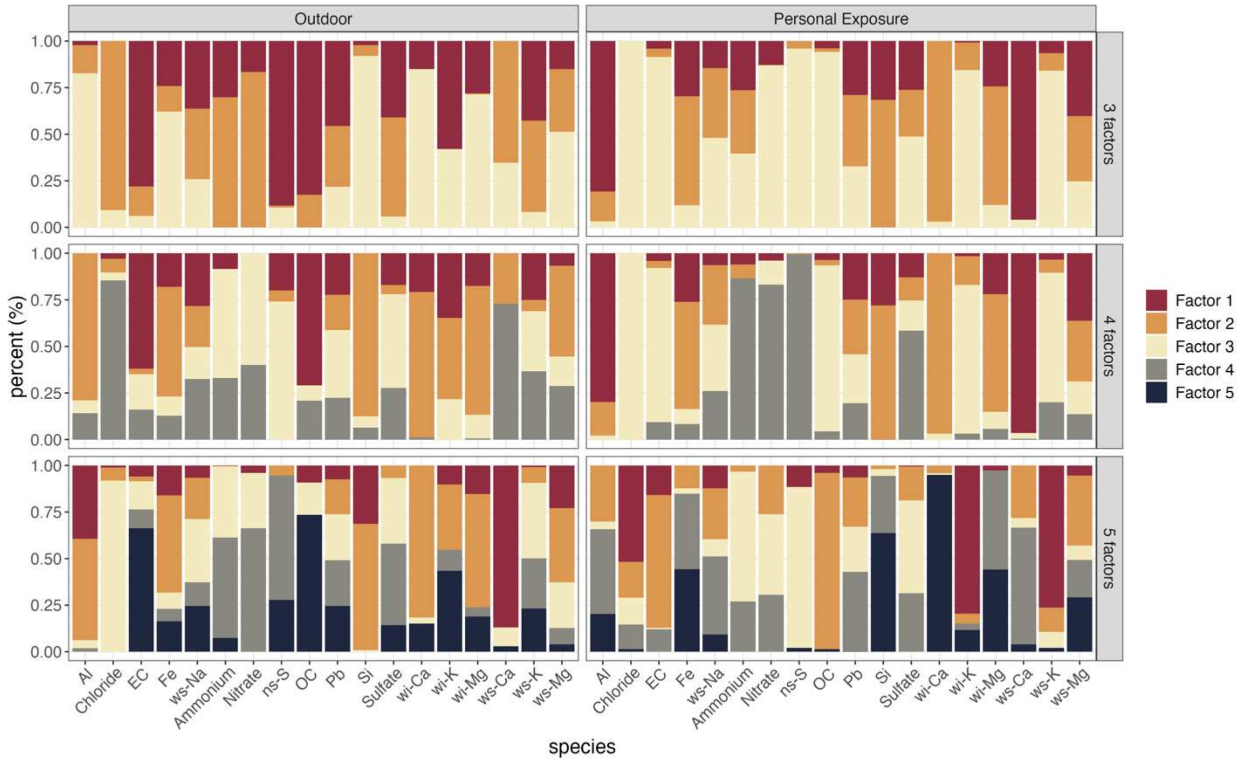
Outdoor



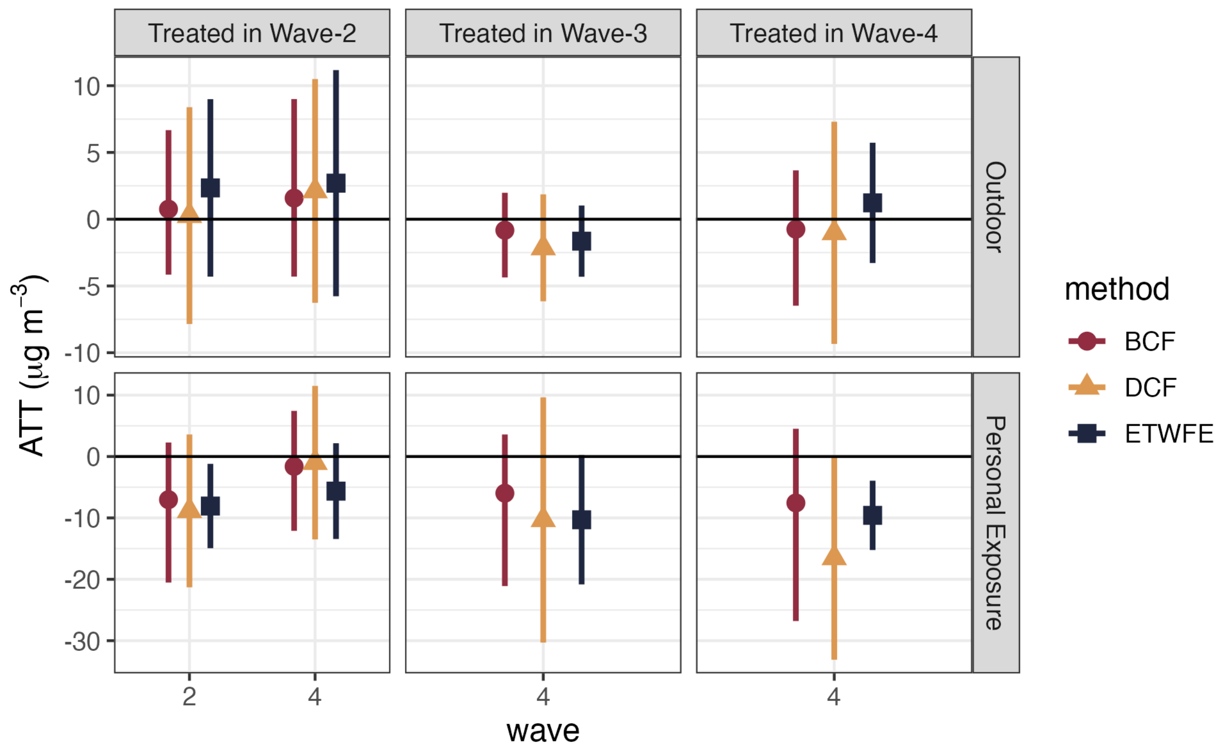
Personal exposure



Appendix C4. Source profiles for the three, four, and give factor source apportionments for individual indoor and personal exposure models.

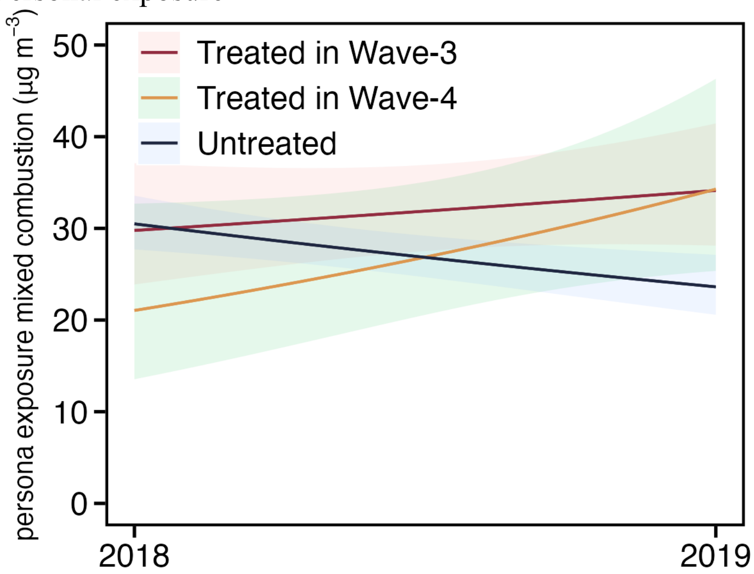


Appendix C5. Time-varying treatment effects for outdoor and personal exposure to the mixed combustion source by treatment cohort estimated by the Bayesian Causal Forests (BCF), Dynamic Causal Forests (DCF), and extended two-way fixed effects model (ETWFE).



Appendix C6. Pre-intervention trends for outdoor and personal exposure to the mixed combustion source.

Personal exposure



Outdoor