NUMBER 235

DECEMBER 2025

Health Effects Institute

RESEARCH REPORT

How Do Household Energy Transitions Work?

Jill Baumgartner, Sam Harper, Chris Barrington-Leigh, Collin Brehmer, Ellison M. Carter, Xiaoying Li, Brian E. Robinson, Guofeng Shen, Talia J. Sternbach, Shu Tao, Kaibing Xue, Wenlu Yuan, Xiang Zhang, and Yuanxun Zhang

INCLUDES A COMMENTARY BY THE INSTITUTE'S REVIEW COMMITTEE

www.healtheffects.org

How Do Household Energy Transitions Work?

Jill Baumgartner, Sam Harper, Chris Barrington-Leigh, Collin Brehmer,
Ellison M. Carter, Xiaoying Li, Brian E. Robinson, Guofeng Shen,
Talia J. Sternbach, Shu Tao, Kaibing Xue, Wenlu Yuan, Xiang Zhang,
and Yuanxun Zhang

with a Commentary by the HEI Review Committee

Research Report 235
Health Effects Institute
Boston, Massachusetts, USA

Trusted Science · Cleaner Air · Better Health

Publishing history: This report was posted at *www.healtheffects.org* in December 2025.

Citation for report:

Baumgartner J, Harper S, Barrington-Leigh C, Brehmer C, Carter EM, Li X, et al. 2025. How Do Household Energy Transitions Work? Research Report 235. Boston, MA: Health Effects Institute.

 Compositor: Progressive Publishing Services, York, PA. Library of Congress Catalog Number for the HEI Report Series: WA 754 R432.

ISSN: 2688-6855 (online)

CONTENTS

ABOUT HEI

The Health Effects Institute is a nonprofit corporation chartered in 1980 as an independent research organization to provide high-quality, impartial, and relevant science on the effects of air pollution on health. To accomplish its mission, the Institute

- identifies the highest-priority areas for health effects research
- competitively funds and oversees research projects
- provides an intensive independent review of HEI-supported studies and related research
- integrates HEI's research results with those of other institutions into broader evaluations
- communicates the results of HEI's research and analyses to public and private decision-makers.

HEI typically receives balanced funding from the US Environmental Protection Agency and the worldwide motor vehicle industry. Frequently, other public and private organizations in the United States and around the world also support major projects or research programs. HEI has funded more than 390 research projects in North America, Europe, Asia, and Latin America, the results of which have informed decisions regarding carbon monoxide, air toxics, nitrogen oxides, diesel exhaust, ozone, particulate matter, and other pollutants. These results have appeared in more than 275 comprehensive reports published by HEI, as well as in more than 2,500 articles in peer-reviewed literature.

HEI's independent Board of Directors consists of leaders in science and policy who are committed to fostering the public–private partnership that is central to the organization. The Research Committee solicits input from HEI sponsors and other stakeholders and works with scientific staff to develop a Five-Year Strategic Plan, select research projects for funding, and oversee their conduct. The Review Committee or Panel, which has no role in selecting or overseeing studies, works with staff to evaluate and interpret the results of funded studies and related research.

All project results and accompanying comments by the Review Committee or Panel are widely disseminated through HEI's website (*www.healtheffects.org*), reports, newsletters, annual conferences, and presentations to legislative bodies and public agencies.

ABOUT THIS REPORT

Research Report 235, *How Do Household Energy Transitions Work?,* presents a research project funded by the Health Effects Institute and conducted by Dr. Jill Baumgartner and Dr. Sam Harper at McGill University, Montréal, Canada, and colleagues. The report contains three main sections:

The **HEI Statement**, prepared by staff at HEI, is a brief, nontechnical summary of the study and its findings; it also briefly describes the Review Committee's comments on the study.

The **Investigators' Report**, prepared by Baumgartner, Harper, and colleagues, describes the scientific background, aims, methods, results, and conclusions of the study.

The **Commentary**, prepared by members of the Review Committee with the assistance of HEI staff, places the study in a broader scientific context, points out its strengths and limitations, and discusses remaining uncertainties and implications of the study's findings for public health and future research.

This report has gone through HEI's rigorous review process. When an HEI-funded study is completed, the investigators submit a draft final report presenting the background and results of the study. Outside technical reviewers first examine this draft report. The report and the reviewers' comments are then evaluated by members of the Review Committee, an independent panel of distinguished scientists who are not involved in selecting or overseeing HEI studies. During the review process, the investigators have an opportunity to exchange comments with the Review Committee and, as necessary, to revise their report. The Commentary reflects the information provided in the final version of the report.

Although this report was produced with partial funding by the United States Environmental Protection Agency under Assistance Award CR–83998101 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and may not necessarily reflect the views of the Agency; thus, no official endorsement by it should be inferred. The contents of this report also have not been reviewed by private party institutions, including those that support the Health Effects Institute, and may not reflect the views or policies of these parties; thus, no endorsement by them should be inferred.

CONTRIBUTORS

RESEARCH COMMITTEE

David A. Savitz, Chair *Professor of Epidemiology, School of Public Health, and Professor of Obstetrics and Gynecology and Pediatrics, Alpert Medical School, Brown University, USA*

Benjamin Barratt *Professor, Environmental Research Group, School of Public Health, Imperial College London, United Kingdom*

David C. Dorman *Professor, Department of Molecular Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, USA*

Christina H. Fuller *Associate Professor, School of Environmental, Civil, Agricultural and Mechanical Engineering, University of Georgia College of Engineering, USA*

Marianne Hatzopoulou *Professor, Civil and Mineral Engineering, University of Toronto, Research Chair in Transport Decarbonization and Air Quality, Canada*

Heather A. Holmes *Associate Professor, Department of Chemical Engineering, University of Utah, USA*

Neil Pearce *Professor of Epidemiology and Biostatistics, London School of Hygiene and Tropical Medicine, United Kingdom*

Evangelia (Evi) Samoli *Professor of Epidemiology and Medical Statistics, Department of Hygiene, Epidemiology and Medical Statistics, School of Medicine, National and Kapodistrian University of Athens, Greece*

Alexandra M. Schmidt *Professor of Biostatistics, School of Population and Global Health, McGill University, Canada*

Neeta Thakur *Associate Professor of Medicine, University of California, San Francisco, USA*

Gregory Wellenius *Beverly A. Brown Professor for the Improvement of Urban Health and Director, Center for Climate and Health, Boston University School of Public Health, USA*

Junfeng Zhang* *Professor of Global and Environmental Health, Duke Global Health Institute*

REVIEW COMMITTEE

Sara D. Adar *Professor of Epidemiology and Global Public Health, Department of Epidemiology, University of Michigan School of Public Health, USA*

Katherine B. Ensor *Noah G. Harding Professor of Statistics, Rice University, USA*

Ulrike Gehring *Associate Professor, Institute for Risk Assessment Sciences, Utrecht University, the Netherlands*

Michael Jerrett *Professor, Department of Environmental Health Sciences, Fielding School of Public Health, University of California, Los Angeles, USA*

Frank Kelly † *Humphrey Battcock Chair in Community Health and Policy and Director of the Environmental Research Group, Imperial College London School of Public Health, United Kingdom*

Jennifer L. Peel † *Professor of Epidemiology, Department of Environmental and Radiological Health Sciences, Colorado State University, and the Colorado School of Public Health, USA*

John Volckens † *Professor, Department of Mechanical Engineering, Walter Scott Jr. College of Engineering, Colorado State University, USA*

---

* *Consultant*

† Review Committee members Kelly, Peel, and Volckens did not participate in the review of this report due to conflicts of interest.

CONTRIBUTORS

HEI PROJECT STAFF

Eleanne van Vliet *Staff Scientist (Study Oversight)*

Hanna Boogaard *Consulting Principal Scientist (Study Oversight)*

Allison P. Patton *Senior Scientist (Report Review)*

Kristin C. Eckles *Senior Editorial Manager*

Tom Zaczkiewicz *Consulting Editor*

PREFACE

HEI's Accountability Research Program

INTRODUCTION

The goal of most air quality actions or interventions is to protect the public health by reducing exposure to air pollutants. If that goal is met and air pollution is reduced, indicators of public health should improve or at least not deteriorate. Evaluating the extent to which air quality actions or interventions succeed in protecting public health is part of a broader effort — variously termed *accountability research, outcomes research, or research on regulatory effectiveness* — designed to assess the performance of environmental policies in general. In recent decades, air quality in the United States and Western Europe has improved substantially, and this improvement is attributable to several factors, including increasingly stringent air quality regulations. However, protecting environmental quality and human health through actions or interventions typically incurs an economic cost. It is, therefore, important to understand whether environmental policies result in the intended improvements.

In 2003, the Health Effects Institute published Communication 11, a comprehensive monograph on accountability research, titled *Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research* (HEI Accountability Working Group 2003). This monograph was written by the members of HEI's multidisciplinary Accountability Working Group after a 2001 National Research Council (NRC) workshop to explore this topic (1998). Communication 11 set out a conceptual framework for accountability research and identified the types of evidence required and the methods by which the evidence should be obtained. It has also guided the development of the HEI Accountability Research program, which is discussed below.

Between 2002 and 2004, HEI issued four requests for applications (RFAs), under which eight studies were funded (see **Preface Table**). A ninth study was funded later, under Request for Preliminary Applications (RFPA) 05-3 Health Effects of Air Pollution. Following this first wave of research, HEI held further workshops to discuss lessons learned, identify key remaining questions, and plan a second wave of research. Those efforts led to further assessments of progress in 2009 and 2010 (HEI 2010a; van Erp and Cohen 2009) and the issuance of *RFA 11-1* Health Outcomes Research — Assessing the Health Outcomes of Air Quality Actions. The first wave of research primarily consisted of studies evaluating relatively short-term, local-scale, and sometimes temporary interventions; RFA 11-1 solicited additional studies with a focus on longer-term, regional- and national-scale regulations, including programs targeted at improving air quality surrounding major ports, as well as further methods development.

This preface describes both the framework of accountability research as it relates to air quality actions or interventions and HEI's Accountability Research program.

BACKGROUND

The first step in assessing the effectiveness of air quality actions or interventions is to measure emissions of the targeted pollutants to see whether they have in fact decreased as intended. To arrive at changes in health that can be attributed to the action or intervention, additional assessments of air quality, exposure, and inhaled dose are needed, as described in detail below. To quantify past effects on health and to predict future effects (US EPA 1999), some accountability studies have used hypothetical scenarios (comparing estimated outcomes under existing and more stringent regulations) and risk estimates obtained from epidemiological studies. However, more extensive validation of those estimates with data on actual outcomes would be helpful.

The long-term improvements in US air quality have been associated with improved health in retrospective epidemiological studies (Chay and Greenstone 2003; Laden et al. 2006; Pope et al. 2009). Considerable challenges, however, are inherent in the assessment of the health effects of air quality actions or interventions. Different actions or interventions occur at different times, for example, and may be implemented at different govermental levels (e.g., national, regional, or local). Therefore, their effectiveness needs to be assessed in ways that take into account the varying times and levels of intervention. In addition, other changes at the same time and place might confound an apparent association between pollution reduction and improved health, such as economic trends (e.g., changes in employment), healthcare improvements, and behavioral changes (e.g., staying indoors when government warnings indicate pollution concentrations are high).

Moreover, adverse health effects that might have been caused by exposure to air pollution can also be caused by other environmental risk factors (some of which might have changed over the same time periods as the air pollution concentrations). These challenges become more pronounced when actions or interventions are implemented over long periods and when changes in air quality and health outcomes are not seen immediately, thus increasing the chance of confounding by other factors. For these reasons, scenarios in which actions or interventions are expected to have resulted in rapid changes in air quality tend to be among the first, and most likely, targets for investigation, rather than evaluations of complex programs implemented over multiple years. Studies in Ireland by Clancy and colleagues (2002) and in Hong Kong by Hedley and colleagues (2002) are examples of such scenarios.

These inherent challenges are well documented in Communication 11 (HEI Accountability Working Group 2003), which was intended to advance the concept of accountability research and to foster the development of methods and studies throughout the relevant scientific and policy communities. In addition, recent advances in data collection and analytic techniques provide an unprecedented opportunity to improve assessments of the effects of air quality interventions.

THE ACCOUNTABILITY EVALUATION CYCLE

Earlier conceptual frameworks for linking air pollution sources to adverse health effects were further developed in HEI's monograph (HEI 2003) in an expanded framework that is still relevant today. This framework can be used to identify factors along an "accountability evaluation cycle" (see **Preface Figure**), each stage of which affords its own opportunities for making quantitative measurements of the intended improvements.

At the first stage (air quality action), one can assess whether controls on source emissions have in fact been put into place. At the second stage (emissions), one can determine whether those controls have indeed reduced emissions, whether emitters have changed their practices, and whether there have been unintended consequences. At the third stage (ambient air quality), one can assess whether reductions in emissions have resulted in improved air quality. At the fourth stage (personal or population exposure), one can assess whether the improvement in air quality has reduced people's actual exposure and whether there has been a benefit for susceptible subpopulations (those most likely to experience adverse health effects). At this stage, it is important to consider changes in time–activity patterns that could either increase or reduce exposure. The actual dose that an individual's organs are exposed to should also be considered (i.e., whether reductions in exposure have led to reductions in concentrations in body tissues such as the lung). Finally, at the fifth stage (human health response), one can assess whether risks to health have declined, given the evidence about changes in health outcomes such as morbidity and mortality that have resulted from changes in exposure. The challenge at this stage is to investigate the health outcomes that are most directly related to exposure to air pollution.

At each stage in the accountability evaluation cycle, the opportunity exists to collect evidence that either validates the assumptions that motivated the intervention or points to ways in which the assumptions were incorrect. The collection of such evidence can thus ensure that future interventions are maximally effective.

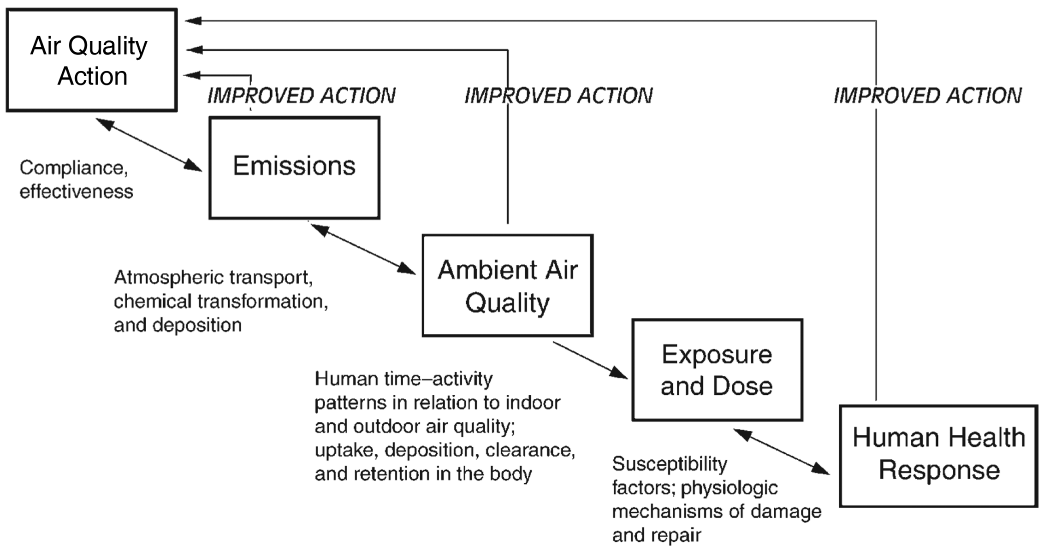
Ultimately, the framework for accountability research will need to encompass investigations of the broader consequences of regulations, not just the intended consequences. Unintended consequences should also be investigated, along with the possibility that risks to public health in fact increased, as discussed by Wiener (1998) and others who have advanced the concept of a portfolio of effects of a regulation.

HEI'S ACCOUNTABILITY RESEARCH PROGRAM

The first wave of HEI's Accountability Research program included nine studies (see Preface Table). These studies involved the measurement of indicators along the entire accountability evaluation cycle, from actions or interventions to human health outcomes. Many of the studies focused on interventions that were implemented over relatively short periods of time, such as a ban on the sale of coal, reductions in the sulfur content of fuels, measures to reduce traffic, and the replacement of old wood stoves with more efficient, cleaner ones. Other studies focused on longer-term, wider-ranging interventions or events; for instance, one study assessed complex changes associated with the reunification of the former East and West Germany, including a switch from brown coal to natural gas for fueling power plants and home-heating systems and an increase in the number of modern diesel-powered vehicles in eastern Germany. HEI also supported research, including the development of methods, in an especially challenging area: assessment of the effects of actions or interventions implemented incrementally over extended periods of time. In one such study, Morgenstern and colleagues (2012) examined changes that resulted from Title IV of the 1990 Clean Air Act Amendments (US EPA 1990), which aimed at reducing sulfur dioxide (SO_2) emissions from power plants by requiring compliance with prescribed emission limitations.

HEI later funded four studies as part of the second wave of its Accountability program (see **Preface Table**). Two studies evaluated regulatory and other actions at the national or regional level implemented over multiple years (Gilliland et al. 2017, Russell et al. 2018); a third study evaluated complex sets of actions targeted at improving air quality in large urban areas and major ports with well-documented air quality problems and programs to address them (Meng et al. 2021); and a fourth study developed methods to support such accountability research (Zigler et al. 2016).

HEI funded a third wave of accountability studies that addressed an array of regional and national actions or interventions (see Preface Table). Sara D. Adar and colleagues evaluated the US EPA's School Bus Retrofit and Replacement Program authorized under the Diesel Emissions Reduction Act. They showed that school attendance and educational achievement



Preface Figure. Accountability evaluation cycle. Each box represents a stage in the process between air quality actions and human health responses to air pollution. Arrows connecting the stages indicate possible directions of influence. The text below the arrows identifies factors affecting the effectiveness of air quality actions at each stage. At several of the stages, knowledge gained from studies on outcomes can provide valuable feedback for improving air quality actions.

had improved in school districts selected for funds to replace old diesel school buses compared to school districts that were not selected for funding (Adar et al. 2024). Perry Hystad and colleagues assessed whether air pollution decreases related to long-term cumulative regulatory improvements of motor vehicle emissions and shorter-term local congestion reduction programs improved birth outcomes among a population of 8.1 million births in Texas between 1996 and 2016 (Hystad et al. 2025). In their Investigators' Report of the current study, Harper and Baumgartner and colleagues examined the impact of a coal heating ban and heat pump subsidy program in Beijing, China, on air quality, air pollutant exposure, and markers of respiratory and cardiovascular health among more than 1,000 participants from an existing cohort. They tested whether effects of the clean heating policy on health might have been because of improvements in wintertime air quality and indoor temperature. Kinney and colleagues investigated sweeping air pollution control policies that began in 2013 across multiple regions of China (in review). They sought to show a causal link between regulations, emissions, ambient air pollution, and mortality over a 10-year period. Funded under a separate RFA, Hakami and colleagues created a source- and location-specific database of mortality benefits per ton of primary $PM_{2.5}$, NO_x, SO_2, and ammonia emissions reductions in the United States and Canada. They showed that emissions reductions in larger cities, particularly primary $PM_{2.5}$, could elicit health benefits nationwide (Hakami et al. 2024).

HEI also continues to fund accountability studies under various other RFAs. A study by Stefanie Ebelt, David Rich, and colleagues was funded under *RFA 20-1A* Health Effects of Air Pollution and is evaluating the effect of selected policies that targeted emissions from motor vehicles and electricity generating units on air quality in Atlanta, New York City, and Los Angeles. Under *RFA 20-1B* Air Pollution, COVID-19, and Human Health Kai Chen of Yale University and colleagues conducted a multicountry study to evaluate whether changes in mortality are associated with changes in ambient NO_2 and $PM_{2.5}$ levels before, during, and after COVID-19 lockdowns in China, Germany, Italy, and the United States (Chen et al. 2024).

Two other accountability-focused studies were funded under the *Walter A. Rosenblith New Investigator Award*. In 2022, Lucas Henneman of George Mason University was funded to estimate the impacts of different emissions sources on daily patterns and concentrations of $PM_{2.5}$ at a fine spatial resolution in the United States. He is performing an accountability analysis of source-related exposure reductions to determine how such reductions have been distributed across the United States. In 2023, a study by Rachel Nethery of Harvard University was funded to develop statistical methods for characterizing variation in health effects across locations and population characteristics associated with exposure to $PM_{2.5}$ across the United States and to design potential policies for reducing $PM_{2.5}$-attributable differences in health outcomes.

A complete list of accountability studies funded by HEI to date is summarized in the Preface Table. The first-wave studies are described in more detail in an interim evaluation of the HEI Accountability Research program (van Erp and Cohen 2009; van Erp et al. 2012). An updated interim discussion of HEI's recent experiences in accountability research is also available (Boogaard et al. 2017).

FUTURE DIRECTIONS

The second and third waves of accountability research were conceived and prioritized during HEI's Strategic Plans for 2010–2015 (HEI 2010b) and 2015–2020 (HEI 2015). In its Strategic Plan for 2020–2025 (HEI 2020a), HEI sought to continue its leadership role in accountability research by prioritizing opportunities for studies that evaluate what methods are best suited to assess the effectiveness of further air-quality improvements. We envision that future studies will again focus on large-scale, complex actions and interventions to improve air quality. We will continue to develop and implement statistical methods, particularly those within a causal inference framework, to tackle these complicated questions. In 2023, HEI issued *RFA 23-2*, which focuses on air quality actions, programs, or other interventions and their effects in local communities that face persistent air pollution challenges in the United States. The selected studies started in 2024.

Throughout its portfolio, HEI emphasizes the importance of data access and transparency because they underpin high-quality research that is used in policy settings. Thus, HEI continues to provide other researchers with access to extensive data and software from HEI-funded studies (see *https://www.healtheffects.org/research/databases*). In the same spirit, the State of Global Air website (HEI 2020b) makes available data on air quality and health outcomes for countries around the world. The interactive site allows exploration of the data and comparisons among countries. The data currently cover 1990–2023 and are updated as new data become available.

REFERENCES

Adar SD, Pedde M, Hirth R, Szpiro A. 2024. Assessing the National Health, Education, and Air Quality Benefits of the United States Environmental Protection Agency's School Bus Rebate Program: A Randomized Controlled Trial Design. Research Report 221. Boston, MA: Health Effects Institute.

Boogaard H, van Erp AM, Walker KD, Shaikh R. 2017. Accountability studies on air pollution and health: The HEI experience. Curr Environ Health Rep; *https://doi.org/10.1007/s40572-017-0161-0*.

Chay KY, Greenstone M. 2003. The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. Q J Econ 118:1121–1167; *https://doi.org/10.1162/00335530360698513*.

Chen K, Ma Y, Marb A, Nobile F, Dubrow R, Stafoggia M, et al. 2025. Effect of Air Pollution Reductions on Mortality During the COVID-19 Lockdowns in Early 2020. Research Report 224. Boston, MA: Health Effects Institute.

Clancy L, Goodman P, Sinclair H, Dockery DW. 2002. Effect of air-pollution control on death rates in Dublin, Ireland: An intervention study. Lancet 360:1210–1214; *https://doi.org/10.1016/S0140-6736(02)11281-5*.

Dockery DW, Rich D, Goodman P, Clancy L, Ohman-Strickland P, George P, Kotlov T. 2013. Effect of Air Pollution Control on Mortality and Hospital Admissions in Ireland. Research Report 176. Boston, MA: Health Effects Institute.

Gilliland F, Avol E, McConnell R, Berhane K, Gauderman WJ, Lurmann FW, et al. 2017. The Effects of Policy-Driven Air Quality Improvements on Children's Respiratory Health. Research Report 190. Boston, MA: Health Effects Institute.

Hakami A, Zhao S, Soltanzadeh M, Vasilakos P, Alhusban A, Oztaner B, et al. 2024. Estimating Model-Based Marginal Societal Health Benefits of Air Pollution Emission Reductions in the United States and Canada. Research Report 218. Boston, MA: Health Effects Institute.

Health Effects Institute (HEI). 2010a. Proceedings of an HEI Workshop on Further Research to Assess the Health Impacts of Actions Taken to Improve Air Quality. Communication 15. Boston, MA: Health Effects Institute.

Health Effects Institute (HEI). 2010b. HEI Strategic Plan for Understanding the Health Effects of Air Pollution 2010–2015. Boston, MA: Health Effects Institute.

Health Effects Institute (HEI). 2020a. HEI Strategic Plan for Understanding the Health Effects of Air Pollution 2020–2025. Boston, MA: Health Effects Institute.

Health Effects Institute (HEI). 2025. State of Global Air 2025: A Report on Air Pollution and Its Role in the World's Leading Causes of Death. Boston, MA: Health Effects Institute.

Health Effects Institute (HEI). 2015. HEI Strategic Plan for Understanding the Health Effects of Air Pollution 2015–2020. Boston, MA: Health Effects Institute.

HEI Accountability Working Group. 2003. Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research. Communication 11. Boston, MA: Health Effects Institute.

Hedley AJ, Wong CM, Thach TQ, Ma S, Lam TH, Anderson HR. 2002. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: An intervention study. Lancet 360:1646–1652; *https://doi.org/10.1016/s0140-6736(02)11612-6*.

Hystad P, Willis M, Hill E, Schrank D, Molitor J, Larkin A, et al. 2025. Impacts of Vehicle Emission Regulations and Local Congestion Policies on Birth Outcomes Associated with Traffic Air Pollution. Research Report 223. Boston, MA: Health Effects Institute.

Kelly F, Anderson HR, Armstrong B, Atkinson R, Barratt B, Beevers S, et al. 2011a. Part 1. Emissions modeling and analysis of air pollution measurements. In: The Impact of the Congestion Charging Scheme on Air Quality in London. Research Report 155. Boston, MA: Health Effects Institute.

Kelly F, Anderson HR, Armstrong B, Atkinson R, Barratt B, Beevers S, et al. 2011b. Part 2. Analysis of the oxidative potential of particulate matter. In: The Impact of the Congestion Charging Scheme on Air Quality in London. Research Report 155. Boston, MA: Health Effects Institute.

Kelly F, Armstrong B, Atkinson R, Anderson HR, Barratt B, Beevers S, et al. 2011c. The London Low Emission Zone Baseline Study. Research Report 163. Boston, MA: Health Effects Institute.

Laden F, Schwartz J, Speizer FE, Dockery DW. 2006. Reduction in the particulate air pollution and mortality: Extended follow-up of the Harvard Six Cities study. Am J Respir Crit Care Med 173:667–672; *https://doi.org/10.1164/rccm.200503-443oc*.

Meng YY, Su JG, Chen X, Molitor J, Yue D, and Jerrett M. 2021. Improvements in Air Quality and Health Outcomes Among

California Medicaid Enrollees Due to Goods Movement Actions. Research Report 205. Boston, MA: Health Effects Institute.

Morgenstern RD, Harrington W, Shih J-S, Bell ML. 2012. Accountability Analysis of Title IV Phase 2 of the 1990 Clean Air Act Amendments. Research Report 168. Boston, MA: Health Effects Institute.

National Research Council (US). 1998. Research Priorities for Airborne Particulate Matter: I. Immediate Priorities and a Long-Range Research Portfolio. Washington, DC: National Academies Press.

Noonan CW, Ward TJ, Navidi W, Sheppard L, Bergauff M, Palmer C. 2011. Assessing the Impact of a Wood Stove Replacement Program on Air Quality and Children's Health. Research Report 162. Boston, MA: Health Effects Institute.

Peel JL, Klein M, Flanders WD, Mulholland JA, Tolbert PE. 2010. Impact of Improved Air Quality During the 1996 Summer Olympic Games in Atlanta on Multiple Cardiovascular and Respiratory Outcomes. Research Report 148. Boston, MA: Health Effects Institute.

Peters A, Breitner S, Cyrys J, Stölzel M, Pitz M, Wölke G, et al. 2009. The Influence of Improved Air Quality on Mortality Risks in Erfurt, Germany. Research Report 137. Boston, MA: Health Effects Institute.

Pope CA 3rd, Ezzati M, Dockery DW. 2009. Fine particulate air pollution and life expectancy in the United States. N Engl J Med 360:376–386; *https://doi.org/10.1056/nejmsa0805646*.

Russell AG, Tolbert PE, Henneman LRF, Abrams J, Liu C, Klein M, et al. 2018. Impacts of Regulations on Air Quality and Emergency Department Visits in the Atlanta Metropolitan Area, 1999–2013. Research Report 195. Boston, MA: Health Effects Institute.

US Environmental Protection Agency (US EPA). 1990. A Bill to Amend the Clean Air Act to Provide for Attainment and Maintenance of Health Protective National Ambient Air Quality Standards, and for Other Purposes. S 1630, 101st Cong, 2nd Sess. US EPA: Washington, DC.

US Environmental Protection Agency (US EPA). 1999. Benefits and Costs of the Clean Air Act 1990 to 2010: Report to Congress. EPA/410/R-99-001. US EPA: Washington, DC: Office of Air and Radiation.

van Erp AM, Cohen AJ. 2009. HEI's Research Program on the Impact of Actions to Improve Air Quality: Interim Evaluation and Future Directions. Communication 14. Boston, MA: Health Effects Institute.

van Erp AM, Cohen AJ, Shaikh R, O'Keefe R. 2012. Recent progress and challenges in assessing the effectiveness of air quality interventions toward improving public health: The HEI experience. EM Magazine October:22–28.

Wiener J. 1998. Managing the iatrogenic risks of risk management. Risk Health Safety Environ 9:39–82. Available: *https://api.semanticscholar.org/CorpusID:70933868*.

Wong C-M, Rabl A, Thach TQ, Chau YK, Chan KP, Cowling BJ, et al. 2012. Impact of the 1990 Hong Kong Legislation for Restriction on Sulfur Content in Fuel. Research Report 170. Boston, MA: Health Effects Institute.

Zhang J, Zhu T, Kipen H, Wang G, Huang W, Rich D, et al. 2013. Cardiorespiratory Biomarker Responses in Healthy Young Adults to Drastic Air Quality Changes Surrounding the 2008 Beijing Olympics. Research Report 174. Boston, MA: Health Effects Institute.

Zigler CM, Kim C, Choirat C, Hansen JB, Wang Y, Hund L, et al. 2016. Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations. Research Report 187. Boston, MA: Health Effects Institute.

Preface Table: HEI's Accountability Research Program

Investigator (Institution)	Intervention	Study or Report Title
	First-Wave Studies[a]	
RFA 02-1		
Douglas Dockery (Harvard T.H. Chan School of Public Health, Boston, MA)	Coal ban in Irish cities	Effect of Air Pollution Control on Mortality and Hospital Admissions in Ireland (Research Report 176; 2013)
Annette Peters (Helmholtz Zentrum München–German Research Center for Environment and Health, Neuherberg, Germany)	Switch from brown coal to natural gas for home heating and power plants, changes in motor vehicle fleet after reunification of Germany	The Influence of Improved Air Quality on Mortality Risks in Erfurt, Germany (Research Report 137; 2009)
RFA 04-1		
Frank Kelly (King's College London, U.K.)	Measures to reduce traffic congestion in the inner city of London	The Impact of the Congestion Charging Scheme on Air Quality in London: Part 1. Emissions Modeling and Analysis of Air Pollution Measurements. Part 2. Analysis of the Oxidative Potential of Particulate Matter (Research Report 155; 2011)
RFA 04-4		
Frank Kelly (King's College London, U.K.)	Measures to exclude most polluting vehicles from entering greater London	The London Low Emission Zone Baseline Study (Research Report 163; 2011)
Richard Morgenstern (Resources for the Future, Washington, DC)	Measures to reduce sulfur emissions from power plants east of the Mississippi River	Accountability Analysis of Title IV Phase 2 of the 1990 Clean Air Act Amendments (Research Report 168; 2012)
Curtis Noonan (University of Montana, Missoula, MT)	Wood stove change-out program	Assessing the Impact of a Wood Stove Replacement Program on Air Quality and Children's Health (Research Report 162; 2011)
Jennifer Peel (Colorado State University, Fort Collins, CO)	Measures to reduce traffic congestion during the Atlanta Olympics	Impact of Improved Air Quality During the 1996 Summer Olympic Games in Atlanta on Multiple Cardiovascular and Respiratory Outcomes (Research Report 148; 2010)
Chit-Ming Wong (University of Hong Kong)	Measures to reduce sulfur content in fuel for motor vehicles and power plants	Impact of the 1990 Hong Kong Legislation for Restriction on Sulfur Content in Fuel (Research Report 170; 2012)
RFPA 05-3		
Junfeng (Jim) Zhang (University of Medicine and Dentistry of New Jersey, Piscataway, NJ)	Measures to improve air quality during the Beijing Olympics	Cardiorespiratory Biomarker Responses in Healthy Young Adults to Drastic Air Quality Changes Surrounding the 2008 Beijing Olympics (Research Report 174; 2013)
	Second-Wave Studies	
RFA 11-1		
Frank Gilliland (University of Southern California)	California and federal programs to improve air quality, including control of emissions from diesel engines and other sources targeted at freight transport and ports, as well as stationary sources	The Effects of Policy-Driven Air Quality Improvements on Children's Respiratory Health (Research Report 190; 2017)

Continues next page

Preface Table (*continued*)

Investigator (Institution)	Intervention	Study or Report Title
Ying-Ying Meng (University of California–Los Angeles)	2006 California Emissions Reduction Plan for Ports and Goods Movement to control emissions from road, rail, and marine transportation, focusing on the ports of Los Angeles and Long Beach	Improvements in Air Quality and Health Outcomes Among California Medicaid Enrollees Due to Goods Movements (Research Report 205; 2021)
Armistead Russell (Georgia Institute of Technology)	Programs to control emissions from major stationary sources and mobile sources in the Southeast United States	Impacts of Emission Changes on Air Quality and Acute Health Effects in the Southeast, 1993–2012 (Research Report 195; 2018)
Corwin Zigler (Harvard T.H. Chan School of Public Health)	National regulations to improve air quality focusing on State Implementation Plans for particulate matter	Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations (Research Report 187; 2016)
	Third-Wave Studies	
RFA 18-1		
Sara D Adar (University of Michigan)	National Clean Diesel Rebate Program in United States	Assessing the national health and educational benefits of the US EPA's school bus retrofit and replacement program: A randomized controlled trial design (In Review)
Sam Harper and Jill Baumgartner (McGill University, Canada)	Coal ban and heat pump subsidy program in the Beijing, China region	How do household energy interventions work? (In Review)
Perry Hystad (Oregon State University)	National and local traffic emissions control measures in Texas	The TRANSIT Accountability Study: Assessing impacts of vehicle emission regulations and local congestion policies on birth outcomes associated with traffic air pollution (In Review)
Patrick L Kinney (Boston University)	Major national air pollution control regulations in China	Accounting for the health benefits of air pollution regulations in China, 2008–2020 (In Review)
RFA 17-2		
Amir Hakami (Carlton University, Canada)	Transportation emission reductions in the United States and Canada	Quantifying Societal Health Benefits of Transportation Emission Reductions in the United States and Canada (Current Report)
RFA 20-1A		
Stefanie Ebelt (Emory University) and David Rich (University of Rochester Medical Center)	Transportation and electricity generation emissions reductions in three US cities	Environmental and Health Benefits of Mobile Source and Electricity Generating Unit Policies to Reduce Particulate Pollution (Ongoing)
RFA 20-1B		
Kai Chen (Yale University)	COVID-19 pandemic lockdown in China, Germany, Italy, and the United States	Effect of air pollution reductions on mortality during the COVID-19 lockdown: A natural experiment study (In Review)
Walter A. Rosenblith New Investigator Award		
Lucas Henneman (George Mason University)	Source-specific emission reductions in the United States	Air pollution source impacts at fine scales for long-term regulatory accountability and environmental justice (Ongoing)
Rachel Nethery (Harvard University)	Health inequity policy design in the United States	Designing optimal policies for reducing air pollution-related health inequities (Ongoing)

[a] Abbreviations: RFA, Request for Applications; RFPA, Request for Preliminary Application.

HEI STATEMENT

Synopsis of Research Report 235

Air Quality and Health Benefits from a Household Clean Heating Policy

BACKGROUND

Accountability research evaluates the extent to which actions or interventions aimed at improving air quality produce the intended reductions in pollutant concentrations and improvements to public health. A major challenge in this research field is isolating changes that can be attributed to the actions in question from improvements that might be due to other unrelated actions or long-term trends. In its 2018 research solicitation, *RFA 18-1* Assessing Improved Air Quality and Health from National, Regional, and Local Air Quality Actions, HEI aimed to fund real-world studies on the effects of air quality actions, including those that were implemented over multiple years or that focused on areas with well-documented air quality problems.

This Statement highlights a study by Jill Baumgartner and Sam Harper (McGill University) and colleagues that was funded under RFA 18-1. The investigators proposed to assess the effects of a household clean heating policy that mandated and subsidized villages in the Beijing region of China to switch from highly polluting residential heaters fueled by coal to efficient electric- or gas-powered heat pumps. They evaluated changes in winter-time air quality and the health of people in 50 villages before any of the villages had the policy and in the first 3 years of staggered policy roll-out in 20 of the villages. They also assessed whether any observed effects of the policy on health could be explained by changes in air quality or indoor temperature in winter.

What This Study Adds

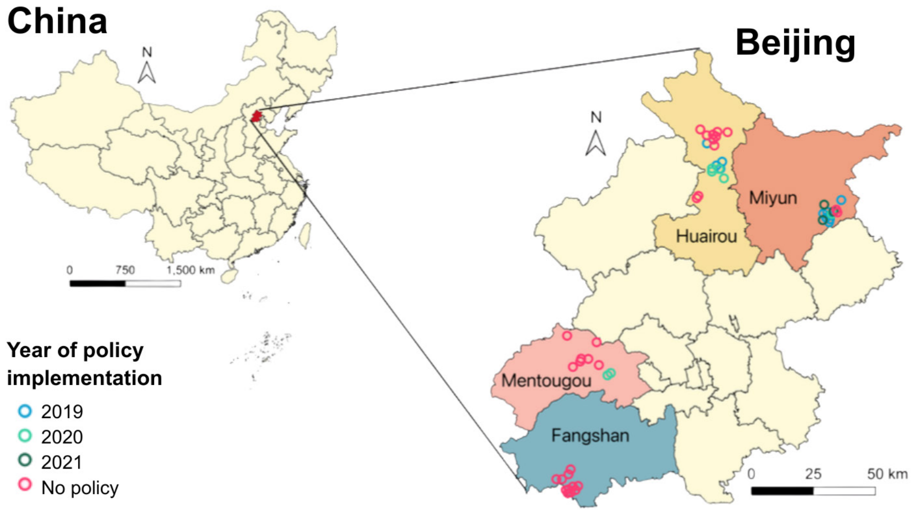
- This accountability study evaluated the effects on air quality and health of a clean heating policy that stipulated and subsidized the conversion of household heating from coal to electric heaters in Beijing-region villages.
- The investigators visited 50 villages in winter and compared changes in air quality, indoor temperatures, and heart and lung health in 20 villages where the policy was implemented with 30 villages where it was not.
- Implementation of the clean heating policy was associated with reduced indoor fine particle concentrations, blood pressure, and respiratory symptoms. The results suggest that the policy's impacts on blood pressure were largely attributable to improvements in air quality and indoor temperature.
- The overall approach was a major strength of this study because it accounted for differences between villages and changes in fuel use, air pollution, and health during the study.
- These results are encouraging for other countries seeking to implement policies to replace highly polluting residential heating sources.

APPROACH

Baumgartner, Harper, and colleagues evaluated changes in 50 Beijing-region villages that were eligible for but not currently participating in the clean heating program in the first year of the study (**Statement Figure**). About half of the villages were expected to enter the program between 2018 and 2022. The clean heating policy was implemented in a subset of those villages each year, in an order determined by various factors related to policy priorities and local capacity. In each village, the investigators recruited about 20 participants living in different households.

The investigators collected data on many parameters in four consecutive winters, starting in late 2018 and finishing in early 2022. They measured fine particles outdoors, indoors in the homes of participants, and with samplers carried by about half of the study participants. They also measured temperature in the participants' homes. For health outcomes, they measured participants' blood pressure and markers of their overall health in blood samples. Additionally, they asked the participants whether they experienced any respiratory symptoms such as shortness of breath or phlegm.

This Statement, prepared by the Health Effects Institute, summarizes a research project funded by HEI and conducted by co-principal investigators Dr. Jill Baumgartner and Dr. Sam Harper at McGill University, Montréal, Canada, and colleagues. Research Report 235 contains the detailed Investigators' Report and a Commentary on the study prepared by the HEI Review Committee.



Statement Figure. Map of village implementation of the clean heating policy. (Source: Adapted from Investigators' Report Figure 1.)

They used sophisticated statistical models to track changes in air pollution and health over time. The models allowed them to compare changes between villages that had implemented the policy each year and those that might implement it in the future. Where they found that the policy led to changes in both air quality and health, they tested whether the changes in health were caused by the changes in air quality or by some other factor related to the policy.

KEY RESULTS

During the study, the investigators evaluated changes in 1,438 participants from 1,236 households in 50 suburban and rural villages of Beijing over four consecutive winters between 2018 and 2022. By the end of the study, the policy had been implemented in 20 of the villages, and there was nearly complete compliance in those villages. Heat pump usage increased from about 2% to about 95% of households in villages where the policy was implemented, and to only about 16% of households where the policy was not implemented. There were corresponding decreases in the amount of coal used by households.

There were some air quality and health improvements in the villages where the policy was implemented. Implementation of the policy reduced indoor fine particle concentrations by about 20 μg/m^3. There were no reductions in outdoor fine particle concentrations and in fine particles measured by samplers carried by participants. At the same time, seasonal average indoor temperatures during winter increased by about 2°C after implementation of the policy. Blood pressure decreased by about 1.5 mm Hg, and respiratory symptoms decreased by about 7.5% when the policy was implemented. Using a statistical approach known as causal mediation analysis, the investigators demonstrated that almost all the effect of the policy on blood pressure could be explained by the improvements in air quality and indoor temperature. In contrast, the improvements in respiratory symptoms could not be explained by these factors and would benefit from further exploration. The policy did not appear to affect the health biomarkers measured in blood.

Because the study included the years of the COVID-19 pandemic, the investigators were forced to make some changes in the measurements and when they were performed. Among the main changes were a partial campaign in winter 2020–2021 and the addition of a fourth full winter data collection campaign (including surveys, air pollution, and health measurements) in winter 2021–2022.

INTERPRETATION AND CONCLUSIONS

In its independent evaluation of the study, the HEI Review Committee thought that Baumgartner, Harper, and colleagues had completed a thorough and important accountability study to evaluate the benefits of a clean energy policy on air quality and health. The Committee identified the overall approach as a major strength of the study because it accounted by design for differences between villages that did not change over time, which helped isolate the effects of changes in fuel use on air pollution levels and health outcomes.

Other explanations for the causes of different exposures and outcomes over time were also considered by the investigators. For example, the fuel use policy was implemented over a period spanning years before, during, and after the COVID-19 pandemic. If the severity of COVID-19 or the community preventative measures against the pandemic, by chance or due to features of the town, were different in the communities with and without the intervention, then bias could occur. Fortunately, this concern was alleviated by the findings that no documented COVID-19 cases were observed in these suburban and rural villages during the study, and that regional restrictions (e.g., travel restrictions, lockdowns, and quarantines) equally affected all the studied communities. They also reported no important differences in pre-intervention trends for personal exposures and health between those groups that did and did not have the clean heating policy, suggesting that their findings were unlikely to be sizably affected by factors that changed over time and could potentially distort the relationship between the policy and its effects. The Committee thought that these and other aspects of the investigators' careful application of complex statistical models to real-world changes contributed to robust main results.

The use of causal mediation analysis was also felt to be a constructive addition. However, the Committee thought that the analyses to understand whether the policy affected health through its effects on air quality and indoor temperature should be treated as exploratory because assumptions had to be made when air pollution and health were not measured at the same time.

Overall, the study demonstrated that the clean heating policy achieved its intended goals to electrify household heating in the villages where it was implemented and that it dramatically reduced residential coal burning and improved indoor environmental quality in the first years after implementation. The policy also provided some benefits to heart and lung health, some of which (blood pressure) appeared to be connected to improvements in air quality.

Although there is an abundance of evidence that air pollution is related to negative health outcomes, it is important to quantify (rather than assume) the effects of specific actions on air quality and health. Here, we have evidence that replacing coal-fueled heaters with heat pumps reduces indoor air pollution and improves health. The results of this study are encouraging for other countries seeking to implement policies to replace highly polluting residential heating sources.

INVESTIGATORS' REPORT

How Do Household Energy Transitions Work?

Jill Baumgartner (co-principal investigator)[1], Sam Harper (co-principal investigator)[1], Chris Barrington-Leigh[1], Collin Brehmer[2], Ellison M. Carter[2], Xiaoying Li[2], Brian E. Robinson[1], Guofeng Shen[3], Talia J. Sternbach[1], Shu Tao[3], Kaibing Xue[4], Wenlu Yuan[1], Xiang Zhang[1], Yuanxun Zhang[4]

[1]McGill University, Montréal, Canada; [2]Colorado State University; Denver, USA; [3]Peking University, Beijing, China; [4]University of the Chinese Academy of Sciences, Beijing, China

ABSTRACT

Introduction Since 2015, thousands of rural and peri-urban villages across Beijing and northern China have been treated by a household Clean Heating Policy (CHP)* that banned household coal burning and subsidized the costs of electric heaters and electricity. Whether this large-scale policy was successful in improving air quality and health remains an important and unresolved question. We estimated the effects of the CHP policy on air quality and cardiopulmonary health in Beijing villages and quantified how much of the policy's effects on health were mediated by changes in air pollution and indoor temperature.

Methods In winter 2018–2019, we enrolled 1,003 participants in 50 Beijing villages that were eligible for, but not currently treated by, the CHP and followed them over four consecutive winter data collection waves. In waves 1, 2, and 4, we administered questionnaires and measured participants' anthropometrics, blood pressure (BP), airway inflammation (fractional concentration of exhaled nitric oxide [FeNO]), and 24-hour personal exposure to fine particulate matter (particulate matter ≤2.5 µm in aerodynamic diameter [$PM_{2.5}$]). Fasting whole blood samples were obtained at clinic visits in waves 1 and 2 for analysis of glucose, lipid profile, and markers of inflammation and oxidative stress. We attempted to contact all prior participants in each follow-up wave. If a previously enrolled participant was not at home or refused subsequent participation, staff first tried to randomly recruit an eligible participant from the same household. If this was not possible, village guides helped field staff to enroll a new participant from a new household using the same sampling procedures as the baseline. Wintertime outdoor $PM_{2.5}$ was measured in all four waves, and wintertime indoor $PM_{2.5}$ was measured in waves 2, 3, and 4. Indoor temperature was measured in all waves. The $PM_{2.5}$ filters were analyzed for their mass, black carbon (BC), and chemical composition, which were used for source apportionment. To estimate the impacts of the policy, we used a difference-in-differences design that accommodated the staggered rollout of the CHP. We used "extended" two-way fixed effects models and marginal effects to quantify the effect of the policy on air pollution and health outcomes. We further evaluated whether villages treated by the policy in different years responded differently to the policy and whether the observed health impacts of the policy were mediated through changes in air pollution or home (indoor) temperature.

Results We enrolled a total of 1,438 participants from 1,236 households during our four study waves. At baseline (wave 1), the mean participant age was 60 years old (standard deviation [SD] = 9.2), 60% of participants were female, and most participants (63%) worked in agriculture. Geometric mean personal exposures to $PM_{2.5}$ were twice as high as outdoor $PM_{2.5}$ (72 vs. 36 µg/m^3), and the main source contributors were local and transported dust, regional and domestic coal and biomass burning, and secondary pollutants. By waves 2, 3, and 4, there were cumulative totals of 10, 17, and 20 villages (of 50 total) exposed to the CHP. Uptake and adherence to the policy were high: among villages treated in wave 2, the proportion of households using heat pumps and coal heaters, respectively, changed from 3% and 97% in wave 1 to 94% and 3% in wave 4, with similar clean energy transitions in villages exposed to the policy in later waves. Marginal effects derived from multivariable extended two-way fixed effects models showed that exposure to the policy increased wintertime indoor temperature by 1° to 2°C and reduced indoor seasonal $PM_{2.5}$ by approximately 20 µg/m^3. Treatment by the policy also reduced contributions to $PM_{2.5}$ from solid fuel sources, including household coal burning, and improved BP (~1.5 mm Hg lower systolic BP [SBP] and diastolic BP [DBP]) and self-reported respiratory symptoms (~8 percentage point reduction in any symptoms). There was notable heterogeneity in effects across treatment cohorts, with larger benefits to indoor $PM_{2.5}$ and health in villages treated in earlier years relative to later years. In the mediation analysis, indoor $PM_{2.5}$

This Investigators' Report is one part of Health Effects Institute Research Report 235, which also includes a Commentary by the Review Committee and an HEI Statement about the research project. Correspondence concerning the Investigators' Report may be addressed to Dr. Jill Baumgartner (email: *jill.baumgartner@mcgill.ca*) and Dr. Sam Harper (email: *sam.harper@mcgill.ca*). No potential conflict of interest was reported by the authors.

Although this report was produced with partial funding by the United States Environmental Protection Agency under Assistance Award CR–83998101 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and may not reflect the views of the Agency; thus, no official endorsement by the Agency should be inferred. The report has also not been reviewed by private party institutions, including those that support the Health Effects Institute, and may not reflect the views or policies of these parties; thus, no endorsement by them should be inferred.

*A list of abbreviations and other terms appears at the end of this volume.

and indoor temperature explained most of the total effect of the policy on SBP and roughly half of the total effect on DBP, but this did not explain improvements in self-reported respiratory symptoms. We did not find evidence of meaningful effects of the policy on outdoor or personal exposure to $PM_{2.5}$ or on biomarkers of inflammation and oxidative stress.

Conclusions In this comprehensive field-based assessment of a large-scale household energy policy in Beijing, we observed high fidelity and compliance with the CHP. Exposure to the policy reduced BP and self-reported chronic respiratory symptoms, and the effects for BP were mediated by reductions in indoor $PM_{2.5}$ and improvements in home temperature, providing empirical evidence that clean household energy policies can provide population health benefits.

INTRODUCTION

China is deploying an ambitious clean energy policy to transition up to 70% of households in its northern provinces from residential coal heaters to gas or electric "clean" space heating, including a large-scale rollout across rural and peri-urban Beijing villages, referred to in this document as the Clean Heating Policy (CHP). To meet this target, the Beijing municipal government announced a two-pronged program that designates coal-restricted areas and simultaneously offers subsidies to nighttime electricity rates and for the purchase and installation of electric-powered heat pumps to replace traditional coal-heating stoves. The policy was piloted in 2015 and, starting in 2016, was rolled out on a village-by-village basis. The variability in when the policy was applied to each village allowed us to treat the rollout of the program as a quasi-randomized intervention and evaluate its impacts on air quality and health. Household air pollution is a well-established risk factor for adverse health outcomes over the entire life course, yet there is no consensus that clean energy interventions can improve these health outcomes based on evidence from randomized trials (Lai et al. 2024). Households may be differentially affected by the CHP because of factors such as financial constraints and user preferences, and there is uncertainty about whether and how the policy may affect indoor and outdoor air pollution, as well as heating behaviors and health outcomes.

BACKGROUND

CONTEXT FOR THE POLICY

The CHP builds on China's long history of launching ambitious, large-scale policies and programs to promote clean household energy transition and support rural energy infrastructure development (Zhang and Smith 2007). China was a relatively early initiator of rural electrification projects in the 1950s and achieved complete (100%) electrification of households by 2016 (Yang 2021), which undoubtedly facilitated the current policy option to replace coal stoves with electric-powered heat pumps. Several decades later, China achieved what is likely still the largest improvement in household energy efficiency in history with regard to the population affected by a single program. The National Improved Stove Program (NISP) and its provincial- and county-level counterparts were initiated in the early 1980s and are credited with introducing 180 million improved cooking and heating stoves by the late 1990s. All NISP stoves had chimneys, and some had manual or electric blowers to promote more efficient combustion (Zhang and Smith 2007), with the primary goal of increased biomass fuel efficiency to promote rural welfare and reduce pressure on local forests and a secondary goal of improving indoor air quality (Sinton et al. 2004). Because NISP focused mainly on biomass cookstoves, it had limited impacts on the rapid increase in coal heating stove installation during that same period, most of which were implemented without chimneys and with rudimentary designs (Zhang and Smith 2007). Although NISP was a significant achievement in the early clean energy transition, especially for biomass cookstoves, the rural energy demands and air pollution challenges of 21st-century China required a renewed effort to promote the transition to cleaner rural energy, particularly for rural heating, where progress significantly lagged behind the energy transition for cooking.

Most of northern China has a temperate continental monsoon climate that is characterized by cold, dry winters and hot, humid summers. Access to central heating is limited to urban areas, and thus most peri-urban and rural households have historically heated their homes using coal heaters and biomass *kangs* (a traditional Chinese energy technology that integrates at least four different home functions, including cooking, a bed for sleeping, space heating, and home ventilation). Household coal burning was a major contributor to indoor and outdoor air pollution in northern China, especially in winter. Before the CHP, more than 100 million rural households consumed approximately 200 million tons of coal to meet more than 80% of northern China's residential space heating demand (Dispersed Coal Management Research Group 2023), which contributed to roughly 30% of wintertime air pollution (GBD MAPS Working Group 2016). In 2013, emissions inventories indicated that coal combustion from industrial, electricity, and residential heating sources was the single largest estimated contributor to population exposures to fine particulate matter (particulate matter ≤2.5 µm in aerodynamic diameter [$PM_{2.5}$]) in China and was responsible for an estimated 366,000 annual premature deaths (GBD MAPS Working Group 2016).

Banning residential coal burning and providing homes with clean heating alternatives through the CHP were considered potentially important interventions to improve rural development, reduce local and regional $PM_{2.5}$, and mitigate air pollution–related health impacts. A number of clean heating options, including electric heat pumps, gas heaters, and electric resistance heaters with thermal storage, were promoted by the Chinese government (Dispersed Coal Management Research Group 2023). By 2021, more than 36 million households in northern China were treated by the CHP, and an estimated 21 million additional households are expected to be treated by 2025. Whether this large-scale energy policy yielded air quality and health benefits remains a critical and unresolved question.

PRIOR EVIDENCE ON HOUSEHOLD ENERGY INTERVENTIONS AND AIR POLLUTION

Household energy interventions, mostly cooking-related ones, that replace traditional solid fuel stoves with more efficient and less polluting alternatives have been implemented and studied extensively in China and other countries over the past several decades. Although the introduction of more efficient household stoves and fuels is expected to reduce indoor air pollution and exposures, evidence of their practical effectiveness in achieving health-relevant air pollution reductions has been mixed, with some studies actually finding worse air quality in homes that received the intervention (Quansah et al. 2017). Furthermore, most previous studies evaluated smaller-scale interventions implemented by civil society organizations or investigators themselves, and the indoor and local air quality benefits of large-scale household energy policies, such as the CHP, have rarely been empirically investigated, especially in countries in the Global South. In Ireland, county-level residential coal bans in the 1990s were associated with 40% to 70% decreases in black smoke concentrations in ban-affected areas (Dockery et al. 2013). In Australia, a wood-burning stove exchange lowered daily wintertime PM_{10} from 44 to 27 μg/m^3 (Johnston et al. 2013), and clean energy policies in New Zealand were associated with 11% to 36% reductions in winter PM_{10} (Scott and Scarrott 2011). The few previous evaluations of the CHP reported small decreases in outdoor $PM_{2.5}$ (–7 to –2.4 μg/m^3) in municipalities or prefectures treated by the policy compared with untreated neighboring regions (Niu et al. 2024; Song et al. 2023; Tan et al. 2023; Yu et al. 2021), and a recent modeling study estimated 36% lower personal exposure to $PM_{2.5}$ based on household-reported changes in fuel use (Meng et al. 2023). These studies captured wide geographic areas, but none included field-based measurements of air pollution or personal exposures, which can differ considerably from modeled estimates (Thompson et al. 2019), and few accounted for secular changes in air quality over time, limiting conclusions about the causal effect of the policy on air quality.

PRIOR EVIDENCE ON CLEAN ENERGY INTERVENTIONS AND CARDIOVASCULAR OUTCOMES

Most previous health assessments of household energy interventions have focused on cookstoves rather than heating technologies, although in many settings, cookstoves are also used for space heating. Randomized trials of less polluting cookstoves generally indicate a cardiovascular benefit. In older Guatemalan women, a chimney stove intervention lowered exposure to air pollution and reduced the occurrence of nonspecific ST-segment depression (McCracken et al. 2011). Randomized trials in Guatemala, Nigeria, and Ghana also showed reductions in blood pressure (BP; SBP range: −3.7 to −1.3 mm Hg) in women assigned to gas, ethanol, or improved combustion biomass stoves. In contrast, recent single country (Peru) and large multicountry (Household Air Pollution Intervention Network [HAPIN]) randomized trials found no benefit of liquefied petroleum gas (LPG) stoves on gestational BP (Checkley et al. 2021; Ye et al. 2022) despite much larger reductions (~66% lower) in exposure to $PM_{2.5}$ and BC than what was observed in trials showing a BP benefit of intervention (Johnson et al. 2022).

The few population-based evaluations of large-scale residential energy policies also suggest a cardiorespiratory benefit of clean energy transition. Residential wood-burning bans were associated with reductions in cardiovascular hospitalizations (–7%) in California (Yap and Garcia 2015) and with reduced cardiovascular (–17.9%) and respiratory (–22.8%) daily mortality in Australia (Johnston et al. 2013), although neither study fully controlled for secular changes in health that were unrelated to the policy. Most relevant to our study are two quasi-experimental assessments of coal replacement policies. In Ireland, reductions in weekly respiratory but not weekly cardiovascular mortality rates were observed following their coal ban (Dockery et al. 2013). A multicity study of Chinese adults in cities where the CHP was piloted compared with adults in cities not in the pilot observed small decreases in chronic lung diseases (–3.0 to –1.1%) but no change in physician-diagnosed cardiovascular diseases, potentially because of the short (1-year) post-policy evaluation period or confounding by other unmeasured municipality-wide air quality or health-related policies (Wen et al. 2023).

Although household air pollution is considered a well-established cause of ill health, the questions of which energy interventions can most effectively reduce air pollution exposures and improve health, and are also scalable and sustainable, remain critical and unanswered. In a recent Official American Thoracic Society Statement, for example, the committee did not reach a consensus on whether household energy interventions (including gas, ethanol, solar, and improved biomass cookstoves) improved health outcomes (including respiratory symptoms and BP), with 55% of the committee saying no and 45% saying yes (Lai et al. 2024).

EVALUATING THE MECHANISMS THROUGH WHICH POLICIES MAY AFFECT HEALTH OUTCOMES

With several notable exceptions (Alexander et al. 2018; Gould et al. 2023; McCracken et al. 2007, 2011), decades of household energy intervention studies have found limited or no health benefit, which demonstrates the complexity of both implementing and evaluating interventions on cooking or space heating that are central to daily life (Ezzati and Baumgartner 2017; Lai et al. 2024). Energy interventions and policies, particularly those implemented at the household or village scales, can produce multiple behavioral, environmental, and health-related changes, making it important to investigate the mechanisms through which such policies exert their health impacts (Dominici et al. 2014). The health benefits achievable with transition from traditional coal stoves to a new electric home heating system, for example, may be influenced by factors that include outdoor air quality (Lai 2020), the desirability and usage patterns of new and traditional stoves (Ezzati and Baumgartner 2017), average or variability in indoor temperature (Lewington et al. 2012), and behaviors such as physical activity or time spent in the home (Lindemann et al. 2017). Only recently were these potential mediating factors considered in health assessments of household energy interventions and rarely in a comprehensive or formal way (Rosenthal et al. 2018). Understanding such mechanisms can provide valuable insights into the success (or failure) of clean energy programs or policies, such as the CHP, in meeting their air quality and health targets and may answer questions that can inform the design of more effective future energy interventions (Lai et al. 2024). For example, is there successful uptake of the policy? Are there cardiovascular-enhancing effects of improved air quality in homes that are treated by the policy? Does the policy lead to heating behavior changes that result in colder homes and thus offset any cardiovascular-enhancing effects of improved air quality? Answers to these questions are facilitated by the analysis of mediating pathways, a key aim of this study.

SPECIFIC AIMS AND OVERARCHING APPROACH

We used three data collection waves in winter 2018–2019, winter 2019–2020, and winter 2021–2022, as well as a partial wave in winter 2020–2021, to advance the following aims:

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

STUDY DESIGN AND METHODS

STUDY AREA

Beijing is the capital of China (population of 21.9 million in 2020) and covers a large geographic area (~16,000 km^2) that includes a highly developed and densely populated urban core surrounded by several satellite towns and thousands of peri-urban and rural villages in the periphery. Beijing winters typically begin in early November and tend to be cold, dry, and windy, with the lowest temperatures most often occurring in January (−3°C, on average), thus requiring space heating (An et al. 2021). Most urban areas of Beijing are connected to a central heating grid that supplies home heating from central locations, whereas rural and many peri-urban areas have historically relied on individual space heating units that, before 2015, were largely fueled by unprocessed coal (Duan et al. 2014).

LOCATION AND PARTICIPANT RECRUITMENT AND ENROLLMENT

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

We recruited approximately 20 households in each village and obtained a household roster for each household. Our tablet-based survey incorporated a randomization tool that randomly ordered household occupants listed on the roster. We recruited a participant in each household by starting at the top of the randomly ordered list until an eligible participant was identified. Household members were eligible to participate if they were older than 40 years old, lived in the study villages, were not planning to move out of the village in the next year, and were not on current immunotherapy or treatment with corticosteroids.

Research staff introduced the study and its measurements to an eligible adult in each household and answered any questions related to the study. In follow-up visits to the study villages, staff first approached households with participants from an earlier wave. Because of study logistics, we were limited to 1 day for study measurements in each village and wave, such that participants who were outside of the village on the measurement day for work or shopping were not able to participate in that wave. If a previous participant was not at home or refused to participate, staff first tried to recruit randomly the next eligible participant listed on the randomized household roster. If there was not another eligible or willing participant in the same household, we recruited a participant from a new household using the same process for household and participant selection described earlier. In wave 1, we recruited 977 new households (**Table 1**). In wave 2, we recruited 189 new households, as well as 81 new participants among the 866 households previously enrolled in Wave 1. In wave 4, we recruited 68 new households, as well as 91 new participants among the 944 existing households recruited in waves 1 and 2. Our village-level study uses individual-level data such that each participant is considered independently.

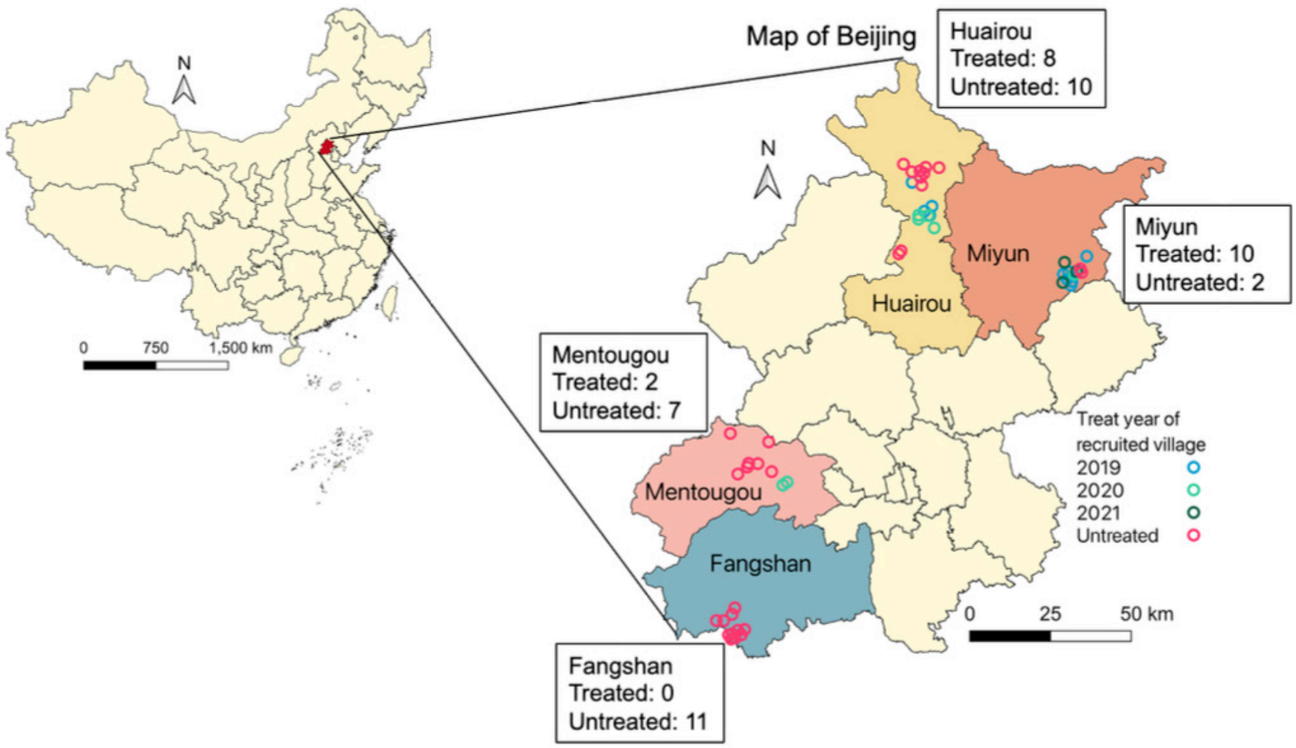


Figure 1. Map of village implementation of the Clean Heating Policy. Each circle represents one recruited village. The colors of the circles indicate the year the villages were exposed to the household energy transition policy.

Table 1. Household Recruitment for Overall and Indoor Air Quality Measurements[a]

	Overall			Indoor			
Sample	**Wave 1**	**Wave 2**	**Wave 4**	**Wave 1**	**Wave 2**	**Wave 3**	**Wave 4**
New recruitment	977	189	68	0	300	0	52
Wave 1 households		866	782		0	0	0
Wave 2 households			162			246	248
Total recruitment	977	1,055	1,012	0	300	246	300

[a] Each column reports the number of households recruited for the listed wave. Each row represents households from the current or previous waves who remained in the study.

All participants provided written informed consent before joining the study. The study protocols were approved by research ethics boards at Peking University (IRB00001052-18090), Peking Union Medical College Hospital (HS-3184), and McGill University (A08-E53-18B).

DATA COLLECTION OVERVIEW

We conducted study measurements over four consecutive waves of data collection in the winter of 2018–2019, 2019–2020, 2020–2021, and 2021–2022 (referred to hereafter as waves 1, 2, 3, and 4, respectively). Field data collection was conducted by approximately 20 trained staff members who traveled to participants' homes to conduct tablet-based household and individual questionnaires, measure participants' BP, and distribute temperature sensors (for measurement of indoor temperature and stove use) and air pollution monitors in all 50 study villages in waves 1, 2, and 4. Anthropometrics (height, weight, and waist circumference), measurement of airway inflammation, and whole blood samples were obtained no more than 1 month later at a village clinic in wave 1 and wave 2. In wave 3, which was during the height of the COVID-19 pandemic, we limited household measurements to indoor air quality and sensor-based measurement of indoor temperature and stove use in 41 villages, including all 17 treated villages and 24 untreated villages, before the Beijing-wide COVID-19–related travel restrictions that halted field data collection. In wave 4, which also occurred during the COVID-19 pandemic, we returned to conducting individual-level assessments. However, unlike in waves 1 and 2, anthropometric measurements and airway inflammation were assessed in participant homes rather than in clinics to avoid group contact, and blood samples were not collected. Outdoor (community) air pollution was measured in all waves.

Air Pollution

Outdoor Air Pollution For outdoor (community) $PM_{2.5}$ monitoring, we deployed between one to three (typically, two) real-time sensors (PMS7003 Plantower, Zefan, Inc.) at different locations in each village. The sensors were assembled with a data logger, electronic screen, and a USB hub into a small metal box that was placed inside an environmental enclosure. One sensor was always placed near the center of the village, and the other one or two sensors were placed no less than 500 meters away from the centrally located sensor. Sensors were positioned at least 1.5 meters above the ground and away from visible point sources of $PM_{2.5}$.

We co-located the real-time sensors with a gravimetric (filter-based) monitor for sensor calibration and analysis of chemical composition for source apportionment. Ultrasonic Personal Aerosol Samplers (UPAS, Access Sensor Technologies, Fort Collins, CO, USA) were used to collect filter-based $PM_{2.5}$ samples with a flow rate of 1.0 L/min (Volckens et al. 2017). The filter-based $PM_{2.5}$ samples were collected and replaced approximately every 7 days throughout the winter, and we rotated the UPAS monitors between villages. Each UPAS was placed inside a custom-built environmental enclosure with a tight fit to prevent any resampling of filtered air. Samplers housed 37-mm polytetrafluoroethylene (PTFE), commonly known as Teflon, filters (VWR, 2.0-μm pore size) and were equipped with a cyclone inlet with a 2.5-μm cut point designed to perform under the sampling flow rate.

In wave 1, we deployed co-located outdoor $PM_{2.5}$ sensors and samplers in 44 of the 50 study villages because of logistical constraints, and we obtained sensor data for 40 villages because of instrument failure in 4 villages. Outdoor data for all 50 villages was obtained in waves 2, 3, and 4. In total, we collected 138, 374, 279, and 295 outdoor $PM_{2.5}$ filter samples in waves 1, 2, 3, and 4, respectively. Field blank PTFE filters were collected at a rate of approximately 10%, subject to the same field conditions as samples.

To support post-sampling determination of organic carbon (OC) and elemental carbon (EC) fractions of $PM_{2.5}$ mass, quartz filters were co-located with a subset of PTFE filter samples collected outdoors. Quartz filter-based $PM_{2.5}$ samples were collected using UPAS operating with a flow rate of 1.0 L/min. UPAS monitors housed 37-mm quartz filters (VWR, 2.0-μm pore size) and were equipped with a cyclone inlet with a 2.5-μm cut point designed to perform under the corresponding sampling flow rate. All quartz fiber filters were baked at 550°C for a minimum of 8 hours to remove organic impurities before sample collection. The $PM_{2.5}$ samples collected on quartz filters were analyzed using established thermo-optical methods for quantifying EC and OC. They were then used to calibrate the colorimetric analysis of EC and OC on PTFE filters (details of this analysis and subsequent calibration are provided under "Optical Properties and Chemical Analysis of $PM_{2.5}$ Mass"). We co-located quartz filters with PTFE filter samples for 23 measurements in wave 2 and 11 measurements in wave 4, along with 3 quartz field blanks in both seasons.

Indoor $PM_{2.5}$ In study waves 2, 3, and 4, we randomly selected six households from the approximately 20 recruited in each village for the measurement of indoor $PM_{2.5}$. In waves 3 and 4, we aimed to monitor indoor $PM_{2.5}$ in the same households sampled in wave 2. If household occupants were not at home or if participants declined indoor $PM_{2.5}$ monitoring, we randomly recruited another household already enrolled in the study. In total, indoor $PM_{2.5}$ was measured in 264 households in wave 2, 346 households in wave 3, and 244 households in wave 4 (Table 1).

Time-resolved indoor $PM_{2.5}$ was measured in all households using the same commercially available sensor (PMS7003 Plantower) used for outdoor sensor-based $PM_{2.5}$ measurements and recorded $PM_{2.5}$ concentrations every 1 minute. The sensor was placed on a table in a room where participants reported spending most of their time (e.g., a living room or bedroom). Indoor $PM_{2.5}$ sensors were deployed between late November and mid-January in each wave, with the start time depending on the village visit date. Measurements continued from the time of deployment until the sensors were collected from homes in late April.

We randomly selected three households from the six with indoor $PM_{2.5}$ measurement to co-locate a filter-based $PM_{2.5}$ sampler. We collected a 24-hour $PM_{2.5}$ filter sample during the first 24 hours of sensor-based measurement. Filter-based $PM_{2.5}$ samples were collected using UPAS or Personal Exposure Monitors (PEMs, Apex Pro) operating with flow rates of 1.0 and 1.8 L/min, respectively. Both samplers housed 37 mm PTFE filters (VWR, 2.0-μm pore size) and were equipped with a cyclone inlet with a 2.5-μm cut point designed to perform under the corresponding sampling flow rate. In total, we collected 150 and 151 indoor $PM_{2.5}$ filter samples in waves 2 and 4, respectively. We did not measure filter-based indoor $PM_{2.5}$ in wave 3 to avoid contact with household occupants during the COVID-19 pandemic. Field blanks were collected at a rate of approximately 10%.

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

Personal Exposure to $PM_{2.5}$ and Black Carbon In waves 1, 2, and 4, we randomly selected approximately 10 study participants in each village for 24-hour personal exposure measurement using two types of $PM_{2.5}$ samplers: PEMs and UPAS. The PEMs actively sampled air at a flow rate of 1.8 L/min, and UPAS sampled air at 1.0 L/min (Volckens et al. 2017). Both samplers housed 37-mm PTFE filters (VWR, 2.0-µm pore size) and were equipped with a cyclone inlet with a 2.5-µm cut point. Sampler flow rates were calibrated the night before deployment and measured immediately after the sampling period. Only 2% of the post-sampling measurements deviated from the target flow rate by more than ±10%. Participants were instructed to wear the sampler in either a small waist pack (for the PEM and sampling pump), an arm band, or a cross-body sling (for the UPAS) for 24 hours, which they could remove from their bodies and place within 2 meters while sleeping, sitting, or bathing. Field blanks for personal air pollution exposure measurements were collected at a rate of approximately 10% in each village. Across waves 1, 2, and 4, study participants contributed 494, 498, and 499 personal $PM_{2.5}$ measurements, respectively.

Gravimetric Analyses of PTFE Filter-Based $PM_{2.5}$ Samples All filters were gravimetrically analyzed (weighed) for their mass before and after deployment at a laboratory at Colorado State University. Briefly, the filters were placed in an environmentally controlled equilibration chamber (21°–22°C, 30%–34% relative humidity) for at least 24 hours before tare and gross weighing. Before each weight was taken, we neutralized static charges by passing the filters over a polonium-210 strip. Filters were weighed on a microbalance (Mettler Toledo Inc., XS3DU, USA) with 1-µg resolution in triplicate or more until the differences among the last three weights were <3 µg. The filters were stored in individually labeled cases and sealed in plastic bags to avoid contamination during transportation and storage. After deployment, filter samples and blanks were immediately stored in a –20°C freezer and, at the end of each field campaign, were transported to Colorado State University, where they were stored in a –20°C freezer before gravimetric and chemical analysis. The difference in the average filter weights from before versus after deployment was used to determine $PM_{2.5}$ mass, which was then blank-corrected using the median mass of blank filters (3 µg for UPAS-collected filters [53% of filter samples]; 33 µg for PEM-collected filters [47% of filter samples]), and $PM_{2.5}$ concentrations were calculated by dividing the mass by the sampled air volume.

We excluded gravimetric (filter) samples meeting any of the following five criteria from the statistical analysis:

1. Run time of less than 80% of a 24-hour target for personal exposure measurement because this is a commonly used cutoff for establishing whether a sample is considered representative of a typical day.
2. Negative mass or extremely high mass values (e.g., >2,000 µg/m^3) that indicate a potential error in data collection or data entry.
3. Missing information on the sampling volume of air.
4. Filters were damaged, including punctures, tears, or holes.
5. Filters were missing during gravimetric analysis and therefore had no mass data.

The final counts of gravimetric $PM_{2.5}$ samples and blanks that met our criteria for inclusion in statistical analysis are given in **Table 2.**

Adjusting Sensor-Based $PM_{2.5}$ Using Filter-Based Gravimetric Measurements Pre- and post-campaign sensor calibrations were conducted to assess whether low-cost sensors responded linearly to $PM_{2.5}$ concentrations measured by co-located federal equivalent method (FEM) instruments. Sensors were deployed alongside the FEM instruments for 7 to 10 days before and after each field campaign. In waves 1, 2, and 3, we co-located the sensors with a rooftop Thermo Electron Synchronized Hybrid Ambient Real-Time Particulate (SHARP) Monitor (model 5030) at Peking University, which is a more urban site. In waves 2, 3, and 4, we additionally co-located the sensors with a rooftop Tapered Element Oscillating Microbalance Method (TEOM, Thermo Scientific 1405 TEOM) at the University of the Chinese Academy of Sciences, which is located in a peri-urban area of Beijing, where we also have study villages. The FEM instrument at Peking University was not functioning after wave 1 data collection, so we instead calibrated the sensors using data from the nearest China National Environmental Monitoring Centre monitor (publicly available at *https://quotsoft.net/air*). The closest distance from the government monitoring stations to Peking University and the Chinese Academy of Sciences University campuses is 1.7 km and 9.9 km, respectively.

We evaluated the performance of all $PM_{2.5}$ sensors for the 7- to 10-day deployments described earlier that occurred before and after each study wave (**Figure 2**). Sensor-measured

Table 2. Count of Total Outdoor and Personal Exposure $PM_{2.5}$ Samples (Filters) Collected Throughout the Project and Number Included in Analyses

	Wave 1		Wave 2		Wave 3		Wave 4	
$PM_{2.5}$ Sample Type	**Total**	**Included[a]**	**Total**	**Included[a]**	**Total**	**Included[a]**	**Total**	**Included[a]**
Outdoor	138	126	374	363	279	213	295	266
Indoor			150	150			151	138
Personal	494	448	498	429			499	418
Blank	52	52	56	56	27	24	101	95

$PM_{2.5}$ = particulate matter ≤2.5 µm in aerodynamic diameter.
[a] Number of samples that met the inclusion criteria for analysis (see text).

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

$PM_{2.5}$ values were highly correlated with values from the FEM instruments (Spearman correlation [ρ] >0.75 in all pre- and post-campaign calibration periods). Daily collections of 24-hour Zeflour (PTFE) and quartz filter samples accompanied the sensors' measurements. For pre- and post-campaign calibration periods, as described earlier, the filter-based data were supplemental to the FEM data, whereas sensor calibration during field campaigns, as described later, was achieved using calibration with concurrently collected filter-based data (because FEM references were not available in the field). We also monitored the sensor-collected data throughout each study wave to identify sensors in need of repair or replacement (e.g., by logging data with the wrong time stamps or only "0" values) and excluded them from deployment. This approach aimed to maintain consistent and accurate measurements from the PM sensors throughout the study.

Sensor calibration during each wave was conducted by deploying filter-based measurements concurrently with sensor-based measurements to establish the linear regression between the low-cost sensor-monitored data and the reference data, and then applying the slope of the linear regression to adjust the low-cost sensor-monitored data. To calibrate the outdoor and indoor measurements of the low-cost sensors in the field, we co-located the sensors with a UPAS (Volckens et al. 2017) or PEM that was used to collect filter-derived $PM_{2.5}$ samples during each field season (wintertime) (Li et al. 2022). The UPAS and PEM were equipped with a cyclone inlet with a 2.5-µm cut point designed to perform under the sampling flow rate of 1 and 1.8 L/min, respectively, and housed a 37-mm PTFE filter (VWR, 2.0-µm pore size). The filter samples were transported to Colorado State University, where they were stored in a −20°C freezer before $PM_{2.5}$ mass measurement.

We established linear regression models between the filter-based $PM_{2.5}$ mass (i.e., the "gold standard" reference) and the sensor-based $PM_{2.5}$ averaged over the same sampling periods. Separate regression models were conducted for indoor and outdoor sensors and each study wave, given the sensitivity of the sensors to relative humidity, temperature, and particle sources, which may differ for indoor versus outdoor conditions and across years. The model slopes were used as the adjustment factors for the sensor-based $PM_{2.5}$ concentrations for that wave. In wave 3, in which only sensor-based measurements were conducted for indoor $PM_{2.5}$, we applied an adjustment factor that was developed from paired indoor filter-sensor data from waves 2 and 4.

We identified larger-than-normal biases and root mean square errors between the sensors and FEM instruments during the post–wave 3 calibration; however, further scrutiny of these data indicated that the differences can be attributed to atypically low air pollution and high humidity during co-location rather than a malfunction of the sensors themselves. Sensor calibration results directly informed the data correction processes applied to account for biases. Generally, the higher correlations between the $PM_{2.5}$ sensor response and FEM data allowed for the application of linear regression models to adjust sensor measurements, and filter-based gravimetric samples were used to validate and correct the sensor data.

We used the adjusted sensor-based $PM_{2.5}$ measurements to calculate a wintertime seasonal mean for indoor and outdoor $PM_{2.5}$ from January 15 to March 15 in each wave to facilitate consistent comparisons across villages in each wave and over time. Additionally, we captured a 24-hour indoor $PM_{2.5}$ concentration that was temporally matched with the timing of personal exposure assessments in the same household to facilitate a comparison between indoor $PM_{2.5}$ and personal exposure results taken during the same period.

Optical Properties and Chemical Analysis of $PM_{2.5}$ Mass We analyzed the optical properties and chemical composition of outdoor and personal exposure gravimetric $PM_{2.5}$ samples to quantify the individual components and species. For each sample, the components were determined by dividing the quantified component mass by the sampled air volume after correcting for field blanks collected in the corresponding study wave.

Following gravimetric analysis, all PTFE filters were analyzed nondestructively for BC using an optical transmissometer data acquisition system (SootScan OT21 Optical Transmissometer; Magee Scientific, Berkeley, CA, USA). Light attenuation through each filter was measured before and after deployment in the field campaign. To calculate BC mass, the difference between the pre- and post-light attenuation was converted to a mass surface loading using the classical Magee mass absorption cross-sections of 16.6 m^2/g for the 880-nm channel optical BC (Ahmed et al. 2009). BC concentrations were calculated by multiplying surface loadings by the sampled surface area of the filters (8.6 cm^2 for UPAS-collected filters; 7.1 cm^2 for PEM-collected filters), correcting for the field blank mass using the median value of blanks (0.31 µg BC for UPAS-collected filters; 0.01 µg BC for PEM-collected filters), and finally dividing by the sampled air volume.

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

The reference measurements for the development of the random forest model of EC and OC were EC and OC measured on quartz filters collected both in study homes and outdoors (as described earlier). The quartz filters were analyzed for OC and EC with a Sunset Laboratory OC/EC Lab instrument (Sunset Laboratories, Inc., MODEL, USA) using the default Sunset Analyzer protocol. A section of each quartz filter underwent a combined thermal desorption-optical transmittance measurement based on NIOSH (National Institute for Occupational Safety and Health) methods 5040 to differentiate and quantify the EC and OC components in $PM_{2.5}$ mass. For the thermal desorption component, the filter is oxidized twice using a strict temperature regime. The first oxidation stage thermally removes OC in a mobile phase of pure helium gas that is converted from carbon dioxide (CO_2) to methane (CH_4) gas and measured by a flame ionization detector (FID). The second oxidation stage proceeds in a mixture of helium and oxygen to oxidize EC, which is also quantified by the FID. The FID is internally calibrated with methane, and external quality control checks are made with sucrose standards. To correct for the potential production of EC during OC pyrolysis in the first oxidation stage, light transmission from a laser through the filter section was monitored throughout the analysis. Reduced light transmittance corresponds to EC generated by the laboratory analysis.

Elemental analysis of $PM_{2.5}$ mass was performed using a Thermo Scientific QUANT'X Evo energy-dispersive X-ray fluorescence (EDXRF) spectrometer with Wintrace software version 10.3 using standard methods (RTI International 2009). Quantitative mass concentrations of 22 individual elements (Mg, Al, Si, S, K, Ca, Ti, Cr, Mn, Fe, Ni, Cu, Zn, Ga, As, Se, Cd, In, Sn, Sb, Te, I) were determined empirically using linear standard curves. Standard curves were generated from commercial, single- and dual-element, thin film standards from MicroMatter Technologies Inc. (Montreal, Canada) in addition to blank films. The quality of the analysis method was evaluated by analyzing a National Institute of Standards and Technology (NIST) standard reference material (SRM) 2783 Air particulate on filter media (Gaithersburg, MD, USA). Elements for which at least 80% of $PM_{2.5}$ mass samples yielded quantifiable element mass were included for positive matrix factorization and source analysis, and apportionment. These elements were Si, Mg, Fe, S, Ca, Al, K, and Pb.

For analysis of water-soluble ions, a portion of each PTFE filter was extracted in 15 mL of deionized water using a Nalgene amber high-density polyethylene bottle and sonicated without heat for 40 min. The extracts were filtered to ensure that insoluble particles were removed using a 0.2-μm PTFE syringe filter. Water-soluble ions were measured using a dual-channel Dionex ICS-3000 ion chromatography (IC) system. Specifically, a Dionex IonPac CS12A analytical (3- × 150-mm) column with eluent of 20-mM methanesulfonic acid at a flow rate of 0.5 mL/min was used to measure cations (Ca^{2+}, Mg^{2+}, Na^+, NH^{4+}, K^+), and a Dionex IonPac AS14A analytical (4 × 250 mm) column with an eluent of 1-mM sodium bicarbonate/8-mM sodium carbonate at a flow rate of 1 mL/min was used to measure anions (SO_4^{2-}, NO^{3-}, Cl^-) (Sullivan et al. 2008).

- ***Water-insoluble species (wi): wi*** refers to the fraction of PM that does not dissolve in water. These species typically include elements such as potassium (K), calcium (Ca), and magnesium (Mg), among others, that remain as PM after water extraction. In this study, because of budget constraints, we did not analyze water-soluble OC. Instead, we determined the water-insoluble fraction by subtracting the amount of the elemental species measured using IC from the total elemental amount determined by X-ray fluorescence (XRF). For instance, the total amount of potassium (K) was measured using XRF, and the water-soluble potassium fraction was quantified using IC. The difference between these two values was taken as the water-insoluble fraction of potassium. This approach is consistent with practices in air quality research that use these analyses.
- ***Water-soluble species (ws): ws*** refers to the fraction of PM that dissolves in water, typically including major ions such as sulfate, nitrate, ammonium, and certain soluble forms of metals. The water-soluble fraction was extracted from particulate samples using deionized water, and the extract was analyzed using IC to determine the concentrations of individual water-soluble ions, such as sulfate (SO_4^{2-}), nitrate (NO^{3-}), and soluble metal ions. This approach is well documented in the scientific literature and follows established protocols for the determination of anions in PM.
- ***Nonsulfate sulfur (ns-S): ns-S*** refers to the sulfur present in PM that is not in the form of sulfate, that is, nonsulfate (ns) sulfur (S). This includes species such as elemental sulfur, organosulfur compounds, and other compounds containing ns-S. The total sulfur content in PM was determined using XRF, consistent with the method we employed for this project (RTI International 2009). The sulfate (SO_4^{2-}) content was quantified using ion chromatography. The ns-S was then calculated by subtracting the sulfate sulfur from the total sulfur determined using XRF analysis. This approach is supported by studies such as those by Shakya and Peltier (2015) and Secrest and colleagues (2016), who have used similar methodologies to distinguish sulfur sources in PM studies.

Outdoor and Indoor (Household) Air Temperature

Hourly outdoor temperature and relative humidity data were obtained from the extensive network of meteorological stations in Beijing. We used digital thermometers (Tianjianhuayi Inc., Beijing, China) to measure indoor "point" temperature in the 5 minutes before BP measurement. Staff measured temperature in a centrally located room, away from heating sources and direct sunlight, by placing the probe in midair at a height that approximated the participant's shoulder height. In a random 75% subsample of households in each wave, we also conducted long-term measurements of indoor temperature by placing a real-time temperature sensor (iButton DS1921G-F5; Thermochron, Maxim Inc., USA) in the room where participants reported spending most of their daytime hours when indoors. Sensors were wall-mounted at a standardized height (~1.5–2 meters), away from major heating sources, windows, and doors, and were programmed to log a temperature reading every 125 minutes for up to 4 months to capture the full winter period and early spring weeks when heating may still intermittently occur. Before the start of each wave, we co-located all of the sensors, measured temperature over 2 days, and compared the readings. Sensors recording values more than 1°C from the group median value were not deployed for data collection.

Objective Measurement of Household Stove Use Using Sensors

Following methods used in a previous intervention evaluation study in rural China (Clark et al. 2017), we objectively measured household heating stove use in a random sample of households selected, also at random, for either short- or long-term measurement. We measured short-term (24-hour) stove use for all household heating stoves in 315 and 227 households in waves 2 and 3, respectively. Long-term stove use was assessed in 324, 273, and 585 homes in waves 2, 3, and 4, respectively, for a period of approximately 6 months. We measured stove use using the same real-time temperature data loggers used to measure seasonal indoor temperature (iButton DS1921G-F5). Field staff placed the sensors on stoves and programmed them to record surface temperature every 125 minutes, a timing decision based on pilot assessments showing that shorter time intervals did not affect the number of heating events detected or heating time recorded. Sensors were placed on the surfaces of biomass and coal-fueled stoves and radiators. For heat pumps, sensors were placed on the heat exchanger coil on air-to-air units and on the radiator of air-to-water units.

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

Questionnaires

Field staff administered household and individual-level questionnaires to assess household demographic information and educational attainment; household assets; house structure; stove and fuel use patterns (including a complete roster of heating methods and their contributions in each room); and individual health behaviors, including exercise frequency, smoking, alcohol consumption, medication use, and clinician-diagnosed health conditions. In wave 4, we recorded actual electricity use in the subsample of households with an available electricity bill on the Wangshangguowang mobile application. We used SurveyBe computer-assisted personal interview software to collect survey data via handheld electronic tablets. Questions were read to participants in Mandarin Chinese, and their responses were recorded into tablets.

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

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

Blood Pressure

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

Self-Reported Respiratory Symptoms and Airway Inflammation

During the questionnaire assessment, participants were asked about chronic airway symptoms, including cough, phlegm, wheeze, and tightness in the chest, using questions validated for use in Mandarin Chinese and developed from the standard St. George's Respiratory Questionnaire (Xu et al. 2009). The Mandarin Chinese questions were extensively piloted with rural and peri-urban Beijing residents to ensure that the health terminology and symptom time patterns were adequate and understandable to the local population.

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

Blood Inflammatory and Oxidative Stress Markers

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

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

We followed standard methods for blood analysis (US FDA 2018). For inflammatory markers (IL-6, TNF-α, CRP), the optical densities (ODs) of all samples were measured using an automated enzyme-linked immunosorbent assay (ELISA) reader. Every plate had eight standard samples used to generate a standard curve that related OD and standard inflammatory marker concentrations. A standard curve for each microplate was generated by a computer software program (BioTek Instruments, Inc., Gen5 Data Analysis Software) based on a four-parameter method. Each plate included at least three control samples to ensure the stability of standard curves. All samples, standards, and controls were measured in duplicate, and the average was used for statistical analysis. For oxidative stress biomarkers (MDA), the chromatographic peak areas of all samples were measured using high-performance liquid chromatography (HPLC) with an ultraviolet detector and HPLC with tandem mass spectrometry. Each plate had seven standard samples used to generate a standard curve that related to the peak area and concentration of the oxidative stress marker. A standard curve for each plate was generated using a computer software program (BioTek Instruments, Inc., Gen5 Data Analysis Software) based on a linear method. Each plate included at least three control samples to ensure the stability of standard curves. Standards and controls for MDA were measured in duplicate, and samples were measured once because of high precision in our pre-analysis pilot study with duplicate testing and evidence from many previous studies showing high stability in measurement (US FDA 2018). Boxplots showing distributions of the inflammatory and oxidative stress markers are provided in Appendix Figure 1.

Anthropometric Measurements

Body weight, height, and waist circumference were measured at the clinic visit in waves 1 and 2 and in participant homes in wave 4 to avoid unnecessary contact during the COVID-19 pandemic. Weight was measured in light indoor clothing without shoes in kilograms to one decimal place using standing scales supported on a steady surface. The scales were calibrated before the start of each wave, and the same staff member weighed themself on the scale each morning to ensure that it was functioning properly. Height was measured without shoes in centimeters to one decimal place with a stadiometer. Waist circumference was measured without clothing obstruction at 1 cm above the participant's navel at minimal respiration in centimeters to one decimal place. The measuring tapes were replaced at the start of each wave to avoid stretching.

MEASURING POLICY IMPACTS

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

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

The canonical DiD design (Card and Krueger 1994) compares two groups (treated and control) at two different time periods (pre- and postintervention, **Figure 3**). In the first period, both groups are untreated, and in the second period, one group is treated (i.e., exposed to the intervention). If we assume that the differences between the groups would have remained constant in the absence of the intervention (the parallel trends assumption), then an unbiased estimate of the impact of the intervention in the post-treatment period can be calculated by subtracting the pre-post difference in the untreated group from the pre-post difference in the treated group. The estimand of interest in a typical DiD analysis is the average treatment effect on the treated (i.e., the *ATT*), which is a contrast of the postintervention outcomes in the treated group with the counterfactual estimate of outcomes in the same population in the absence of treatment.

When multiple groups are treated at different times, as with the CHP, the most common approach has been to use

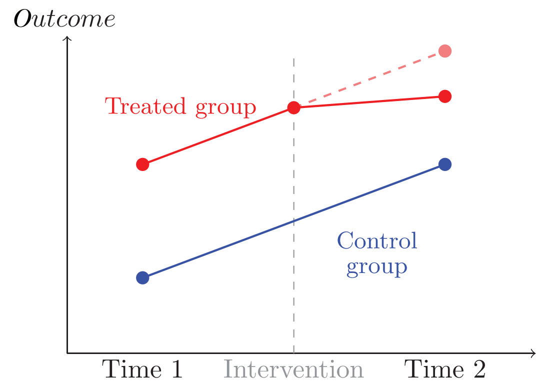
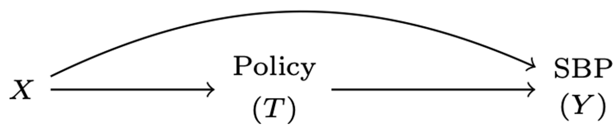


Figure 3. Stylized example of difference-in-differences.

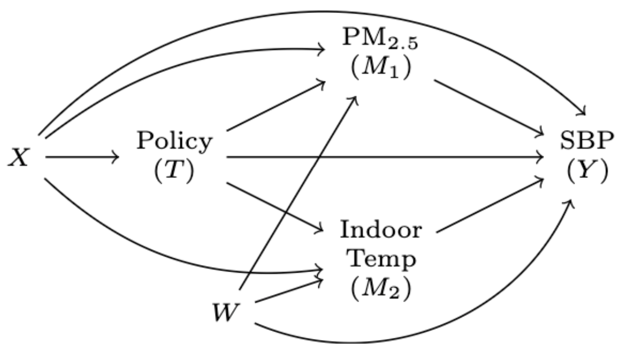
a two-way fixed effects model to estimate the impact of the intervention, which controls for secular trends and fixed differences between villages. However, recent evidence suggests that traditional two-way fixed effects estimation of the treatment effect may be biased in the context of heterogeneous treatment effects (i.e., where the effects of treatment vary for different groups treated at different times; Callaway and Sant'Anna 2021; Goodman-Bacon 2021). The bias occurs because when there are multiple groups treated at different times, the two-way fixed effects estimate is a weighted average of several "2 × 2" DiD estimates, some of which involve using already treated units as controls for later treated units (Baker et al. 2022). We take advantage of new developments in the econometrics literature (Callaway and Sant'Anna 2021; Sun and Abraham 2021; Wooldridge 2021) that relax the assumption of homogeneity in the context of staggered policy rollouts, but also allow straightforward interpretation of *ATT*s for assessing policy impacts. This decision was motivated by the many behavioral, social, or economic factors that might affect both new heat pump use and coal stove suspension (e.g., energy prices and availability, wintertime temperature, COVID-19 pandemic, user preferences) over time in our study, and thus the possibility that the effect of the policy on air pollution and health may be dynamic over time or heterogeneous across treatment cohorts.

MEASURING PATHWAYS AND MECHANISMS

To estimate how much of the CHP may work through different mechanisms, we used causal mediation analysis. Causal approaches to mediation attempt to discern between and clarify the necessary assumptions for identifying different kinds of mediated effects. **Figure 4** shows directed acyclic graphs (DAGs) to illustrate (a) the total effect and (b) the potential direct and indirect effects of the CHP, with T as the policy, X as a set of pretreatment covariates, and Y SBP as an example outcome (DAGs for other outcomes are given in the Appendix). The "total effect" of the policy is an estimate of how much the overall outcomes (Y) would change for a



(a) DAG for total effect



(b) DAG for direct and indirect effects

Figure 4. Hypothetical directed acyclic graphs (DAGs) showing (a) total effect and (b) direct and indirect effects with outcome (*Y*), pretreatment covariates (*X*), policy (*T*), and multiple mediators (M_1,M_2), as well as covariates for the mediators (*W*).

change in exposure (vs. no exposure) to the CHP (*T*). Part (b) of Figure 4 adds M_1 as $PM_{2.5}$ and M_2 as indoor temperature to represent potential mediators that are affected by the policy. In this scenario we can estimate the controlled direct effect (*CDE*), which is the effect of the CHP on BP had we intervened to set M_1 and M_2 to fixed levels (i.e., the *CDE* represents the effect of the CHP that goes through pathways other than the mediators). For example, we can estimate the impact of the policy on health outcomes while holding $PM_{2.5}$ and indoor temperature at uniform levels of average background exposure or some other hypothetical level (VanderWeele 2015).

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

DATA ANALYSIS

To understand how the policy's impact on health may be mediated by different potential mediators, we need to first estimate the total effect of the policy on the outcomes shown in Figure 4 (a) and then estimate the *CDE*s after adjustment for potential mediators and any residual mediator-outcome confounding. As previously discussed, for the mediators to "explain" the total effects of the policy on health, the policy should affect the mediators, and the mediators should also affect the outcomes.

TOTAL EFFECT

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

$$Y_{ijt} = g\left(\mu_{ijt}\right) = \alpha + \sum_{r=q}^{T} \beta_r \boldsymbol{d}_r + \sum_{s=r}^{T} \gamma_s \boldsymbol{fs}_t + \sum_{r=q}^{T}\sum_{s=r}^{T} \tau_{rs}\left(\boldsymbol{d}_r \times \boldsymbol{fs}_t\right) + \varepsilon_{ijt} \tag{1}$$

in which Y_{ijt} is the outcome for individual *i* in village *j* at time *t*, d_r represent treatment cohort dummies (i.e., fixed effects for each cohort of villages *r* that were first exposed to the policy at the same time *q*, e.g., in 2019, 2020, or 2021), fs_t are fixed effects for each time period *s* corresponding to different winter data collection waves (2018–2019, 2019–2020, or 2021–2022), and τ_{rs} are the treatment cohort-time *ATT*s in the context of a linear model. For nonlinear outcomes, the cohort-time *ATT*s are derived by estimating average marginal effects from nonlinear models (Arel-Bundock 2024). For binary and count outcomes, we used logit and Poisson models, respectively, and for skewed outcomes (e.g., $PM_{2.5}$, BC, inflammatory markers), we used generalized linear models with a gamma distribution and a log link based on the specification tests recommended by Manning and Mullahy (2001). For all models, we clustered standard errors at the village level, consistent with the unit of treatment assignment (Cameron and Miller 2015). The ETWFE and other approaches that allow for several (potentially heterogeneous) treatment effects may also be averaged to provide a weighted summary *ATT*. Several potential possibilities are feasible, including weighting by treatment cohorts or time since policy adoption (Goin and Riddell 2023). We generally focus on two types of *ATT*s for this report: simple averages across all treatment cohorts and the full set of cohort-time *ATT*s to evaluate heterogeneous treatment effects. Although we primarily focus on reporting the simple average *ATT* for most outcomes, we also used omnibus joint *F*-tests to assess whether there was sufficient evidence to reject the assumption of homogeneity across the *ATT*s. Although we conducted hypothesis tests to evaluate heterogeneity, in general, we focus on estimation and confidence intervals rather than null hypothesis significance testing in interpreting model estimates. We refrain from categorizing results as "significant" based on arbitrary alpha thresholds, or basing conclusions solely on whether p-values cross a specific threshold, consistent with recommended statistical practice of the American Statistical Association (Wasserstein et al. 2019).

MEDIATION ANALYSIS

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

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

where δ is now the conditional effect of the mediator M at the reference level of the treatment (again, represented via the series of group-time interaction terms), and the collection of η_{rs} terms are coefficients for the product terms allowing for mediator-treatment interaction. Finally, ζ is a vector of coefficients for the set of confounders contained within W. As noted earlier, in the staggered DiD framework that allows for heterogeneity, we do not have a single treatment effect but a collection of group-time treatment effects that may be averaged in different ways. This extends to the estimation of the *CDE*, in which case we will also have several *CDE*s that can be averaged to make inferences about the extent to which the policy's impact is mediated by $PM_{2.5}$ or temperature. Based on the setup in Equation 2, the *CDE* is estimated as $\tau_{rs} + \eta_{rs}M$. In the absence of interaction between the exposure and the mediator (i.e., $\eta_{rs} = 0$) the *CDE* will simply be the estimated treatment effects $\sum_{r=q}^{T}\sum_{s=r}^{T}\tau_{rs}$ (i.e., the effect of the policy holding M constant). For a valid estimate of the *CDE*, we must account for confounding of the mediator-outcome effect, represented by W in the equation earlier. The inclusion of baseline measures of both the outcome and the proposed mediators inherent in our DiD strategy help to reduce the potential for unmeasured confounding of the mediator-outcome effect (Keele et al. 2015). Given the large number of outcomes of interest in this study, as well as the potential for heterogeneous treatment effects, we limited the mediation analysis to health outcomes for which we observed some evidence of a total effect of the CHP.

IDENTIFICATION OF POTENTIAL CONFOUNDERS AND MODEL COVARIATES

In contrast to typical analytic approaches such as regression adjustment or propensity scores that solely focus on measured covariates, our DiD approach helps to minimize the risk of some sources of *unmeasured* confounding. Treatment cohort fixed effects control for measured and unmeasured time-constant factors that may differ between treatment cohorts (e.g., genetics, altitude), and time fixed effects control for secular trends, capturing any unmeasured factors that affect outcomes in all treatment cohorts (including the untreated) similarly over the study period (e.g., background improvements in ambient air quality or household transition to more efficient heating unrelated to the CHP). The latter are particularly helpful in the context of the documented declines in ambient $PM_{2.5}$ in the Beijing region attributable to air quality improvement programs and policies other than the CHP policy (van Donkelaar et al. 2021; Zhang et al. 2019).

For models estimating the effect of the policy on indoor temperature and health outcomes, we used DAGs (Pearl 2000) to identify potential time-varying causes of both treatment by the policy and our study outcome(s) that could differ between treatment groups and adjusted for those potential confounders in the regression models. For the mediation analysis, we identified potential mediator-outcome confounders using the same approach. These variables were identified from the relevant peer-reviewed literature and our team's substantive knowledge about the CHP. For models estimating the effect of the policy on air pollution outcomes, the main predictors of personal exposures and indoor air quality in rural China are inconsistent across studies (e.g., Lee et al. 2021; Ni et al. 2016). Thus, we considered the following covariates as potential determinants of air pollution in our study setting: village population and total number of households in the village; outdoor temperature, relative humidity, dew point, wind direction, wind speed, and boundary layer height; home area and home area heated; home insulation; active or passive smoking status of the participant; and whether the household reported residential wood (i.e., biomass) burning, and if so, self-reported quantity used.

Exposure to tobacco smoke is important for both air pollution and health outcomes, and we used the participant responses to survey questions related to their current smoking status (i.e., is the participant a current smoker, former smoker, or never smoker) and, for never smokers, history of passive smoking (i.e., has the participant ever lived with a smoker in the same house for at least 6 months, with possible responses of never, yes but not currently, and yes at present). The survey responses were used to create the following four distinct tobacco smoking categories: *current smoker*, defined as currently smoking at the time of survey; *former smoker*, defined as previously smoking but no longer smoking at the time of the survey (not accounting for duration of cessation); *never smoker with a history of living with a smoker*, defined as a never smoker who is currently living with a smoker or has previously lived with a smoker for at least 6 months (i.e., passive smoking exposure); and *never smoker*, defined as no history of smoking and no history of living with a smoker for more than 6 months.

Ultimately, we included the following measured time-varying covariates in the final DiD models for each outcome-specific model. Models for air pollution outcomes were adjusted for household size, tobacco smoking category, outdoor temperature, and outdoor dew point. As a sensitivity analysis, we additionally adjusted for district of residence, given the baseline district-level differences in energy use, socioeconomic status, and altitude, especially for villages in Fangshan compared with the other three districts. Temperature models were adjusted for the number of rooms, wintertime occupants in the household, age of the primary respondent, and wealth index. Models for BP were adjusted for age, sex, waist circumference, tobacco smoking category, alcohol consumption, and use of BP medication. For self-reported respiratory outcomes, we adjusted for age, sex, tobacco smoking category, occupation, frequency of drinking, and frequency of farming. Measured respiratory outcome (FeNO) models included adjustment for age, sex, body mass index (BMI), frequency of drinking, tobacco smoking category, frequency of exercise, occupation, and time of measurement. Inflammatory marker outcome models were adjusted for age, waist circumference, occupation, wealth index quantile, frequency of drinking, tobacco smoking category, and frequency of farming. For the final covariate-adjusted DiD model for personal exposure "mixed combustion" source contributions, we adjusted for temperature (represented by a spline with 2 degrees of freedom), tobacco smoking category, and whether the household reported using biomass fuel. For the final covariate-adjusted DiD model for outdoor (community) "mixed combustion" source contributions, the following covariates were included: total number of households in the village, village population, and ambient relative humidity (represented by a spline with 2 degrees of freedom).

MULTIPLE IMPUTATION FOR COVARIATES AND INDOOR $PM_{2.5}$ IN ANALYSES WITH BLOOD PRESSURE OUTCOMES

Blood pressure was measured at household visits, but several key covariates — such as waist circumference, height, and weight — were measured at the clinic visits in waves 1 and 2. In wave 4, blood pressure and anthropometrics were all measured during household visits. Thus, we were missing covariate information for individuals in waves 1 and 2 who were unable to attend the clinic visits (~15%–20% of participants in each wave). Additionally, because we only measured indoor $PM_{2.5}$ in a randomly selected subsample of 300 homes in waves 2 and 4, we were missing indoor $PM_{2.5}$ for all participants in wave 1 with BP measures, as well as for a subsample of participants in waves 2 and 4. To prepare data for the BP outcomes analysis, we used multiple imputation with chained equations (MICE) to impute missing indoor $PM_{2.5}$ and missing covariate data values for individuals who participated in the household visit but not the clinic visit. This allowed us to retain observations with BP measurements that would have otherwise been dropped in adjusted and mediation models using complete-case analysis. Imputation was performed with the *MICE* package (van Buuren and Groothuis-Oudshoorn 2011) in *R* (m = 30 imputation datasets, with 30 iterations each), and the DiD and mediation analyses were conducted for each of the 30 datasets. We then used Rubin's rules to combine point estimates and standard errors while accounting for both within- and between-dataset variances (Rubin 1987).

Appendix Table 1 shows that most measures had no or less than 1% missing data, except for measured waist circumference, height, and weight (all ~15% missing). Appendix Table 2 also shows the number and percentage of missing observations by treatment enrollment cohort and outcome, and we found little evidence that the percentage of missing observations differed substantially between treatment cohorts. In Appendix Figure 2, we show kernel density plots for the distribution of imputed values for BMI, waist circumference, and indoor $PM_{2.5}$, all of which closely approximated the observed values.

RESULTS

We retained all 50 study villages during this 4-year longitudinal assessment of village treatment by the CHP, although we were only able to visit 41 villages in winter 2020–2021 (wave 3), when we were limited to village and household-level measurements of air quality, indoor temperature, and stove use because of travel restrictions during the COVID-19 pandemic.

By waves 2, 3, and 4, there were cumulative totals of 10, 17, and 20 (out of 50 total) study villages treated by the CHP policy, respectively. All of the treated villages in our study selected to install electric-powered air-source heat pumps with 200 RMB (Renminbi, also known as the Chinese yuan) per m^2 (up to 24,000 RMB) in subsidies and were also provided with 80% nighttime electricity subsidies up to 10,000 kWh per heating season. To limit coal use, villages enrolled in the policy were no longer allowed to place orders for subsidized coal with the district-level governments that manage the procurement and distribution of coal for residential heating in Beijing. In addition, village committee leaders in treated villages reported feeling accountable to the Environmental Protection Department for limited coal-related air pollution and were motivated to encourage residents not to burn coal. Some villages were equipped with government air pollution monitors, and the Environmental Protection Department conducted village inspections and issued warnings about coal burning. Households burning coal in treated villages were at risk of losing their electricity subsidy.

Appendix Figure 8 and Appendix Table 3 show the participation of villages, households, and participants across the four waves of data collection and the number of sampled participants, households, and villages by study wave. Appendix Table 4 also shows selected demographic characteristics by district. We conducted measurements in more than 1,000 participants in each of the three measurement waves that included individual-level measurements. **Table 3** shows selected demographic characteristics and health behaviors between participants who contributed to each study wave. In total, **Table 4** shows that we enrolled 1,438 participants into the study, of which 630 (43%) individuals contributed 1,890 observations across all three waves in which health

Table 3. Selected Demographic and Health Characteristics of Participants in Each Study Wave

	Estimates			Test for Equality	
Characteristic	**Wave 1 (2018–2019) (*n* = 1,003)**	**Wave 2 (2019–2020) (*n* = 1,110)**	**Wave 4 (2021–2022) (*n* = 1,028)**	**Statistic[a]**	***P* Value**
Female, *n* (%)	597 (59.5)	654 (58.9)	617 (60.0)	0.270	0.874
Current smoker, *n* (%)	257 (25.6)	295 (26.6)	265 (25.8)	0.292	0.864
Passive smoke exposure, *n* (%)	486 (48.4)	538 (48.5)	486 (47.3)	0.234	0.890
Any smoke exposure, *n* (%)	795 (79.3)	898 (80.9)	857 (83.4)	5.616	0.060
Age in years, mean (SD)	60.1 (9.3)	61.1 (9.1)	63.3 (9.0)	31.980	0.000
Body mass index (kg/m^2), mean (SD)	26.1 (3.7)	25.7 (3.5)	26.2 (3.8)	4.209	0.030
Waist circumference (cm), mean (SD)	86.8 (10.2)	87.4 (9.4)	91.3 (10.4)	54.171	0.000

SD = standard deviation.
[a] Chi-square test for categorical and F-test for continuous characteristics.

Table 4. Selected Demographic and Health Characteristics of Participants Who Contributed to Different Numbers of Study Waves

	Estimates			Test for Equality	
Characteristic	**1 Wave** $n_{participants} = 365$ ($n_{obs} = 365$)	**2 Waves** $n_{participants} = 443$ ($n_{obs} = 886$)	**3 Waves** $n_{participants} = 630$ ($n_{obs} = 1{,}890$)	**Statistic[a]**	***P* Value**
Female, *n* (%)	211 (57.8)	532 (60.0)	1,125 (59.5)	0.542	0.763
Current smoker, *n* (%)	110 (30.1)	230 (26.0)	477 (25.2)	3.817	0.148
Passive smoke exposure, *n* (%)	172 (47.1)	425 (48.0)	913 (48.3)	0.099	0.952
Any smoke exposure, n (%)	293 (80.2)	732 (82.6)	1,526 (80.7)	1.607	0.448
Age in years, mean (SD)	26.3 (3.6)	25.8 (3.6)	26.0 (3.7)	1.532	0.433
BMI (kg/m^2), mean (SD)	59.8 (9.3)	61.0 (8.9)	62.1 (9.3)	11.718	0.000
Waist circumference (cm), mean (SD)	90.3 (9.8)	88.1 (10.2)	88.7 (10.2)	4.438	0.024

BMI = body mass index; SD = standard deviation.
[a] Chi-square test for categorical and F-test for continuous characteristics.

measurements were conducted, 443 (31%) individuals contributed 886 observations across two waves, and 365 (25%) individuals participated in a single wave. We found no differences in sex or smoking across waves, but overall BMI and waist circumference increased over time. Table 4 shows similar demographic and health characteristics of participants who contributed to a different number of waves, with some evidence that individuals contributing to more than one wave of data had slightly higher BMI and smaller waist circumference.

DESCRIPTION OF THE STUDY SAMPLE BY TREATMENT

Table 5 shows the distribution of selected demographic, health, and environmental characteristics from the baseline survey before any villages were enrolled in the CHP. We provide means and standard deviations separately for villages that eventually enter into the policy and those that never do so. As noted earlier, although our DiD identification strategy allows for fixed differences between treated and untreated

Table 5. Descriptive Statistics for Selected Demographic, Health, and Environmental Measures at Baseline by Treatment Status

	Never Enrolled (*n* = 603)		Ever Enrolled (*n* = 400)			
	Mean	**SD**	**Mean**	**SD**	**Difference in Means**	**SE**
Demographics						
Age (years)	59.9	9.4	60.4	9.2	0.5	0.6
Female (%)	59.5	49.1	60.0	49.1	0.5	3.2
No education (%)	11.5	31.9	12.3	32.9	0.9	2.1
Primary education (%)	75.5	43.0	77.6	41.7	2.1	2.8
Secondary+ education (%)	12.6	33.2	9.8	29.7	–2.9	2.0
Wealth index (bottom 25%)	26.9	44.4	22.3	41.7	–4.6	2.8
Wealth index (25%–50%)	23.6	42.5	27.0	44.5	3.4	2.9
Wealth index (50%–75%)	24.7	43.1	25.5	43.6	0.8	2.9
Wealth index (top 25%)	24.8	43.2	25.2	43.5	0.4	2.9
Health Measures						
Never smoker (%)	21.8	41.3	19.1	39.4	–2.7	2.6
Former smoker (%)	11.9	32.4	15.1	35.8	3.2	2.2
Passive smoker (%)	39.6	49.0	40.2	49.1	0.6	3.2
Current smoker (%)	26.2	44.0	25.4	43.6	–0.8	2.8
Never drinker (%)	55.9	49.7	52.5	50.0	–3.4	3.2
Occasional drinker (%)	26.0	43.9	25.5	43.6	–0.5	2.8
Daily drinker (%)	17.8	38.3	21.9	41.4	4.1	2.6
Systolic (mm Hg)	131.4	16.8	128.7	14.3	–2.7	1.0
Diastolic (mm Hg)	82.7	11.6	82.1	11.3	–0.6	0.8
Waist circumference (cm)	87.7	10.5	85.4	9.5	–2.3	0.8
BMI (kg/m^2)	26.3	3.7	25.8	3.6	–0.5	0.3
Frequency of coughing (%)	18.7	39.0	19.7	39.8	1.0	2.6
Frequency of phlegm (%)	27.6	44.7	23.7	42.6	–3.8	2.8
Frequency of wheezing (%)	6.2	24.2	6.6	24.8	0.3	1.6
Shortness of breath (%)	29.2	45.5	34.3	47.5	5.1	3.0
Chest trouble (%)	11.6	32.0	14.1	34.9	2.5	2.2
Any respiratory problem (%)	50.6	50.0	54.3	49.9	3.7	3.2
Environmental Measures						
Temperature (°C)	13.8	3.6	13.5	3.3	–0.3	0.2
Personal $PM_{2.5}$ (μg/m^3)	127.1	145.3	102.3	105.5	–24.7	11.9
Black carbon (ug/m^3)	4.4	5.3	3.3	3.4	–1.1	0.4

BMI = body mass index; $PM_{2.5}$ = particulate matter ≤2.5 μm in aerodynamic diameter; SD = standard deviation; SE = standard error.

villages, overall, the differences at baseline are generally small, and the groups seem well balanced on most measures, except personal exposure to $PM_{2.5}$, which was lower in villages that were eventually treated.

SUMMARY OF PARTICULATE MATTER AND BLACK CARBON MEASUREMENTS

At baseline, before the policy was rolled out in any study villages, $PM_{2.5}$ and BC concentrations were higher, on average, for personal exposures compared with outdoor concentrations. From wave 2 onward, with the inclusion of indoor air pollution measurements, personal exposure air pollution concentrations were still higher than indoor or outdoor concentrations, with indoor levels being higher than outdoors (**Table 6**). This trend (personal > indoor > outdoor) was observed among households in treated and untreated villages. Personal, indoor, and outdoor geometric mean (GM) (95% confidence interval [CI]) concentrations of $PM_{2.5}$ were 72 (65, 80), 45 (39, 53), and 33 (29, 36) μg/m^3, respectively, and elevated relative to health-based guidelines. The current World Health Organization (WHO) guidelines state that annual average exposures to $PM_{2.5}$ should not exceed 5 μg/m^3, and 24-hour average exposures should not exceed 15 μg/m^3 for more than 3 to 4 days per year (WHO 2021). Interim targets have been set to support the planning of incremental milestones toward cleaner air, particularly for cities, regions, and countries with higher air pollution levels. For $PM_{2.5}$, the four interim (IT) targets for annual and 24-hour means are IT-1: 35 and 75 μg/m^3; IT-2: 25 and 50 μg/m^3; IT-3: 15 and 37.5 μg/m^3; and IT-4: 10 and 25 μg/m^3 (WHO 2021). The baseline personal exposures to $PM_{2.5}$ in our study aligned with IT-1,

Table 6. Arithmetic and Geometric Means for Air Pollutant Concentrations (micrograms per cubic meter) by Wave

			Wave 1		Wave 2		Wave 3		Wave 4	
			Est.	CI	Est.	CI	Est.	CI	Est.	CI
Personal Exposure Measurements										
Filter-derived	24-h $PM_{2.5}$	Mean	117	[105, 129]	97	[87, 107]			84	[72, 96]
		GM	72	[65, 80]	60	[54, 66]			47	[42, 52]
	24-h BC	Mean	3.9	[3.5, 4.4]	3.6	[2.9, 4.2]			3.7	[2.9, 4.5]
		GM	2.6	[2.4, 2.8]	1.9	[1.7, 2.1]			1.7	[1.5, 1.9]
Indoor Measurements										
Sensor-derived	Seasonal $PM_{2.5}$	Mean			94	[84, 103]	84	[75, 94]	67	[59, 75]
		GM			71	[65, 77]	63	[57, 70]	47	[42, 52]
Filter-derived	24-h $PM_{2.5}$	Mean			69	[59, 78]			58	[48, 68]
		GM			45	[39, 53]			33	[27, 40]
	24-h BC	Mean			2.7	[2.1, 3.2]			2.9	[2.2, 3.5]
		GM			1.6	[1.3, 2.0]			1.6	[1.3, 1.9]
Outdoor Measurements										
Sensor-derived	Seasonal $PM_{2.5}$	Mean	47	[45, 48]	55	[54, 56]	33	[32, 34]	33	[32, 34]
		GM	36	[35, 37]	40	[39, 41]	23	[22, 23]	22	[22, 23]
Filter-derived	Seasonal $PM_{2.5}$	Mean	38	[34, 42]	38	[34, 41]	25	[23, 28]	26	[24, 28]
		GM	33	[29, 36]	30	[28, 32]	21	[19, 23]	22	[21, 24]
	Seasonal BC	Mean	1.5	[1.3, 1.6]	1.4	[1.3, 1.5]			1.2	[1.1, 1.2]
		GM	1.3	[1.1, 1.4]	1.1	[1.0, 1.2]			1.0	[0.9, 1.1]

BC = black carbon; CI = confidence interval; Est. = estimated; GM = geometric mean; $PM_{2.5}$ = particulate matter ≤2.5 μm in aerodynamic diameter.

indicating considerable opportunity for air quality exposure reduction with intervention.

We also present the geometric and arithmetic means (and 95% CIs) for $PM_{2.5}$ and BC in each measurement wave (Table 6). Wave 3 (2020–2021) was a partial wave that took place over a period affected by the COVID-19 pandemic and did not involve filter-based air pollution sample collection.

POLICY UPTAKE

Each year of the study, participants reported the types of fuels and stoves and the amount of fuel used for space heating in winter. Based on these data, heating energy types were classified into four categories: exclusive use of a heat pump ("Heat pump exclusively"), use of a heat pump and a biomass-fueled kang ("Heat pump with biomass kang"), use of solid fuel heater with electric heating devices other than heat pumps ("Coal stove and biomass kang with other electric heater"), and exclusive use of solid fuel ("Coal stove and biomass kang"). In villages treated by the policy, **Figure 5** shows meaningful transitions from solid fuel to electric-powered heat pumps for all treatment cohorts. For example, the proportion of households in the group treated in 2019 (wave 2) reporting any use of heat pumps increased from 3% in wave 1 to 93% in wave 2 and 96% in wave 4. Conversely, any use of coal stoves decreased from 97% in wave 1 to 8% in wave 2 and 3% in wave 4. We observed similar stove-use transitions for households in villages treated in 2020 (wave 3). In the three villages treated in 2021, we observed less exclusive use of the heat pump and a slightly larger proportion of households continuing to use coal.

We also observed a substantial decline in the amount of self-reported coal used in villages treated by the CHP (**Figure 6**), although the reduction in coal use was smaller with each subsequent treatment cohort (Appendix Table 5). Biomass (i.e., wood logs or twigs, charcoal), usually burned in kangs for both cooking and space heating, was not expressly targeted by the CHP. We observed declines in self-reported biomass use in villages treated in 2019 and 2020, but there was a small increase in biomass consumption in the cohort treated last (2021).

In never-treated villages, we also observed a transition from solid fuel to clean energy over the 4-year study, but it was much slower than in villages exposed to the CHP. The proportion of households that reported using electric heat pumps increased from 5% in wave 1 to 10% in wave 2 and 25% in wave 4, and those who adopted heat pumps tended to use them exclusively. Commensurately, the reported expenditures on electricity increased gradually over time in the untreated villages. The percentage of untreated households using solid fuel with other types of electric devices remained relatively stable, ranging from 64% to 70% across waves. Self-reported use of biomass also remained stable, at approximately 1 ton of fuel each winter, whereas exclusive use of solid fuel decreased from 30% in wave 1 to 7% in wave 4.

To evaluate potential bias in self-reporting, we compared self-reported versus actual winter electricity use in a subsample (n = 37) of households with available electricity bills. Both untreated and treated households incorrectly reported their winter electricity expenses, although treated households were particularly prone to underestimating electricity use compared with untreated households (Appendix Figure 9).

AIM 1: POLICY IMPACTS AND POTENTIAL MEDIATION

Impact of Policy on Potential Mediators of Air Pollution and Indoor Temperature

The *ATT* from the basic ETWFE model (**Table 7**) shows that exposure to the CHP reduced 24-hour indoor $PM_{2.5}$ by –18.9 µg/m³ (95% CI: –56.1, 18.4). After adjusting for outdoor temperature, outdoor dew point, smoking category, and the number of residents in each household, the *ATT* was –20.0

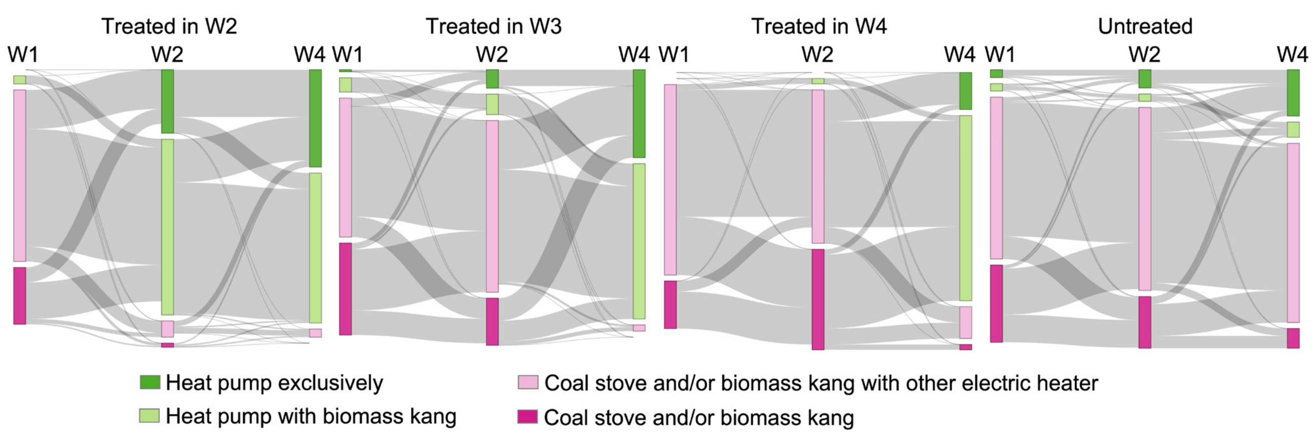


Figure 5. Transitions to different energy sources across study waves 1, 2, and 4.

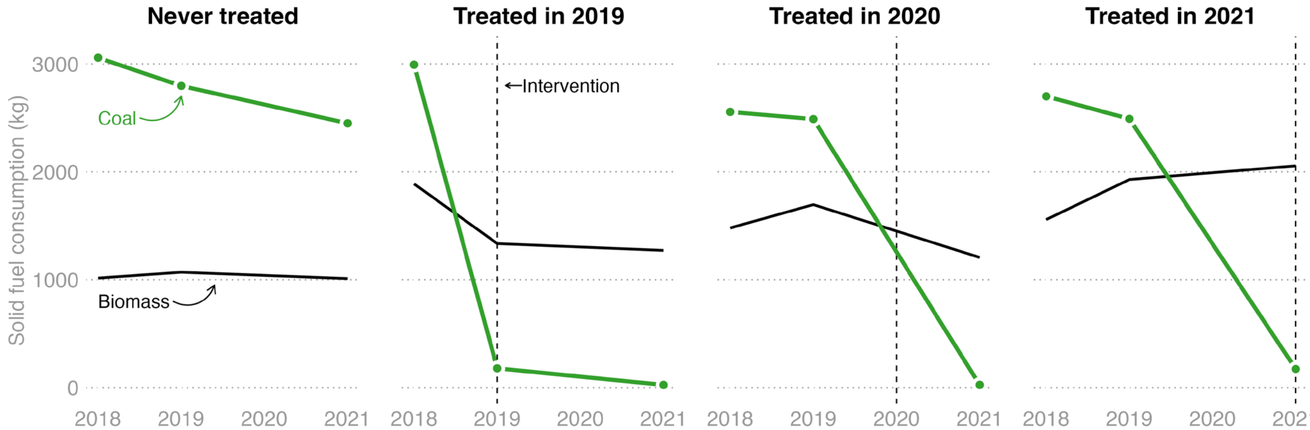


Figure 6. Trends in self-reported coal and biomass consumption by treatment season.

Table 7. Treatment Effect on Outdoor and Indoor $PM_{2.5}$, Personal Exposure to $PM_{2.5}$ and Black Carbon, and Measures of Indoor Temperature[a]

			DiD		Adjusted DiD	
		Obs	***ATT***	**(95% CI)**	***ATT***[b]	**(95% CI)**
			Air pollution (µg/m³)			
Personal	$PM_{2.5}$	1,270	−3.0	(−26.1, 20.1)	0.2	(−19.6, 19.9)
	Black carbon	1,161	−0.6	(−1.7, 0.6)	−0.4	(−1.5, 0.6)
Indoor	24-h $PM_{2.5}$[c]	399	−18.9	(−56.1, 18.4)	−20.0	(−45.6, 5.5)
	Seasonal $PM_{2.5}$	366	−30.9	(−53.2, −8.7)	−20.3	(−37.5, −3.0)
Outdoor	24-h $PM_{2.5}$	11,174	−0.5	(−5.5, 4.4)	−2.1	(−10.0, 5.8)
	Seasonal $PM_{2.5}$	139	1.7	(−3.4, 6.7)	0.5	(−4.8, 5.9)
			Indoor temperature (°C)			
Point	Mean	2,999	1.9	(0.9, 2.9)	1.9	(0.9, 2.9)
Seasonal	Mean (all)	1,350	0.7	(−0.1, 1.4)	0.7	(−0.1, 1.4)
	Mean (daytime)	1,346	0.8	(0.0, 1.5)	0.8	(0.0, 1.5)
	Mean (heating season)	1,350	1.8	(0.9, 2.7)	1.8	(0.9, 2.7)
	Mean (daytime heating season)	1,346	2.0	(1.0, 2.9)	1.9	(1.0, 2.9)
	Minimum (all)	1,350	4.2	(2.3, 6.0)	4.2	(2.4, 6.1)
	Minimum (heating season)	1,350	4.2	(2.3, 6.0)	4.2	(2.4, 6.0)

ATT = average treatment effect on the treated; DiD = difference-in-differences; ETWFE = extended two-way fixed effects; $PM_{2.5}$ = particulate matter ≤2.5 µm in aerodynamic diameter.

[a] Outdoor and indoor $PM_{2.5}$ were derived from sensor measurements after being adjusted based on co-located gravimetric $PM_{2.5}$ measurements. 24 h indicates the mean $PM_{2.5}$ concentrations during the 24 hours when personal exposure samples were collected in each village. "Seasonal" indicates the seasonal mean $PM_{2.5}$ concentrations in each village, from January 15th to March 15th.

[b] ETWFE models for air pollution outcomes were adjusted for household size, smoking category, outdoor temperature, and outdoor dew point. Temperature models adjusted for the number of rooms and wintertime occupants in the household, age of the primary respondent, and wealth index.

[c] The indoor 24-h $PM_{2.5}$ concentration was determined over the time period concurrent with when the personal $PM_{2.5}$ concentration was determined.

μg/m³ (95% CI, –45.6, 5.5). The basic DiD impact was stronger on seasonal indoor $PM_{2.5}$, with an average *ATT* of –30.9 μg/m³ (95% CI: –53.2, –8.7). After adjustment, the average *ATT* on seasonal indoor $PM_{2.5}$ was –20.3 μg/m³ (95% CI: –37.5, –3.0). This finding likely reflects the direct benefit of the policy in replacing coal stoves and improving air quality. We found little evidence of heterogeneity in *ATT*s across cohort and time (all *P* values >0.4 for tests of heterogeneity; see Appendix Table 7). Overall, we found little evidence of an impact of the CHP on different measures of outdoor (local, community-level) $PM_{2.5}$ or personal exposures to $PM_{2.5}$ and BC. Adjusted *ATT*s were –2.1 μg/m³ and 0.5 μg/m³ for 24-hour and seasonal outdoor $PM_{2.5}$, respectively. Adjusted estimates for personal $PM_{2.5}$ and personal BC were 0.2 μg/m³ and –0.4 μg/m³, respectively, and generally, all of the estimates for outdoor and personal exposure impacts were imprecise (Table 7). We provide the full set of cohort-time *ATT*s for personal exposure in Appendix Table 6. Appendix Tables 19 and 20 show the impact of including wave 3 data on the estimates of the impact of the policy on indoor (seasonal) and outdoor (24-hour and seasonal) $PM_{2.5}$, respectively. Generally, this improved precision but did not affect the magnitude of our estimates. Further adjustment for district of residence in the covariate-adjusted models had little impact on our results (Appendix Table 17).

With respect to the other potential mediator of the effect of the CHP, temperature, Table 7 shows that exposure to the CHP increased mean household point temperature by 1.9°C (95% CI: 0.9, 2.9) in the adjusted DiD model, with similar impacts on mean seasonal temperatures during the heating season. The CHP had considerably stronger impacts on average seasonal minimum temperatures, which increased by 4.2°C (95% CI: 2.4, 6.0) in the adjusted DiD model. Notably, we observed little difference in the ATTs between the basic DiD and the models additionally adjusting for the number of rooms and wintertime occupants in the household, age of the primary respondent, and wealth index.

Impact of the Policy on Health Outcomes

Table 8 shows the impacts of the policy on BP in basic ETWFE models and models further adjusted for age, sex, waist circumference, smoking category, alcohol consumption, and use of BP medication. Overall exposure to the CHP demonstrated reductions in BP of approximately 1.5 mm Hg for both SBP and DBP in adjusted DiD models, but we found little evidence of a meaningful impact on pulse pressure (PP) or BP amplification. The effects of the policy on brachial and central BPs were similar and were consistent when restricted to only those participants enrolled in wave 1 (i.e., excluding participants recruited in later waves; see Appendix Table 18). However, the average effects in Table 8 conceal a fair amount of heterogeneity in treatment effects for BP across treatment cohorts and time. Appendix Table 9 shows that treatment impacts were considerably stronger for the earlier compared to the later-treated cohorts. For example, the CHP reduced central DBP in the year of treatment by –2.7 mm Hg (95% CI: –4.6, –0.8) for the villages first treated in wave 2, but increased central DBP by 1.1 mm Hg (95% CI: –0.1, 2.2) for the villages treated in wave 4 (*P* value for heterogeneity <0.0001).

Table 8 shows the impacts on self-reported chronic respiratory symptoms categorized as any symptoms and separately for each symptom type. Based on the covariate-adjusted ETWFE models, exposure to the CHP reduced self-report of any poor respiratory symptoms by 7.5 percentage points (95% CI: –12.7, –2.3). This overall effect was mostly due to reductions of roughly 3 percentage points in reports of coughing, having chest trouble, or difficulty breathing, respectively, on several or most days of the week. We found limited evidence that the CHP reduced self-reported symptoms of phlegm (–1.6, 95% CI: –5.6, 2.4) or wheezing (1.0, 95% CI: –1.9, 3.9). Appendix Tables 11–16 show limited evidence of systematic heterogeneity in the cohort-time treatment effects across the outcomes of any symptom, coughing, or trouble breathing, but we did observe heterogeneity in the overall *ATTs* for phlegm (Appendix Table 13), wheezing (Appendix Table 14), and chest trouble (Appendix Table 16). More specifically, we observed increases in phlegm and wheezing symptoms and stronger decreases in chest trouble in later cohorts treated in 2020 and 2021. This heterogeneity may explain the reduced overall *ATTs* for these outcomes compared with the relatively larger group-time effects.

Table 8 also shows the impacts of the CHP on FeNO, which was conducted in a subsample of 511 participants, including 274 participants with one measurement, 142 with two measurements, and 95 participants with three measurements. Of these, 22 participants with 50 observations were missing covariate data, and thus, the adjusted DiD was based on 793 observations from 489 participants. We found little evidence that the policy affected changes in FeNO in the covariate-adjusted ETWFE model (0.3 ppb, 95% CI: –2.2, 2.8). There was some evidence of heterogeneity in the FeNO effects of the policy by treatment cohort Appendix Figure 10, although the CIs for each of the cohort-specific effects were wide and overlapping. Our results did not substantively change with sensitivity analyses that limited the analysis to participants with at least two repeated measurements (ATT: –0.6, 95% CI: –3.2, 2.0) and to those who participated in all three waves (adjusted *ATT*: 0.3, 95% CI; –2.9, 3.4), where all estimates remained close to zero and were statistically consistent with the adjusted DiD in the main analysis (Appendix Table 21).

We also found limited evidence of an impact of the CHP on markers of inflammation or oxidative stress. The basic DiD analyses showed an increase of 1.0 (95% CI: 0.1, 1.8) in TNF-α, but this estimate was reduced to 0.8 (–0.1, 1.7) after adjustment for waist circumference, occupation, wealth index quantile, frequency of drinking, smoking category, and frequency of farming. The adjusted estimates for the effect of the policy on IL-6, CRP, and MDA were also generally small and measured with limited precision.

Table 8. Overall Impacts of the CHP on Blood Pressure, Respiratory Outcomes, Inflammatory Markers, and Malondialdehyde

			DiD		Adjusted DiD	
		Obs	***ATT***	**(95% CI)**	***ATT*[a]**	**(95% CI)**
BP[a]						
Systolic BP (mm Hg)	Brachial	3,082	−0.8	(−2.6, 1.0)	−1.4	(−3.3, 0.5)
	Central	3,081	−1.0	(−2.8, 0.7)	−1.6	(−3.4, 0.3)
Diastolic BP (mm Hg)	Brachial	3,082	−1.3	(−2.6, 0.0)	−1.6	(−3.0, −0.2)
	Central	3,081	−1.4	(−2.7, −0.0)	−1.7	(−3.0, −0.3)
Pulse pressure (mm Hg)	Brachial	3,082	0.5	(−0.7, 1.7)	0.2	(−1.0, 1.4)
	Central	3,081	0.3	(−0.8, 1.5)	0.1	(−1.0, 1.2)
BP amplification ×100	Pulse pressure	3,081	0.1	(−1.1, 1.4)	−0.0	(−1.2, 1.2)
	Systolic BP	3,081	0.2	(−0.2, 0.5)	0.1	(−0.2, 0.4)
Respiratory Outcomes[b,c]						
Self-reported (pp)[b]	Any symptom	3,076	−7.7	(−12.8, −2.5)	−7.5	(−12.7, −2.3)
	Coughing	3,076	−2.6	(−7.2, 2.0)	−2.7	(−7.1, 1.7)
	Phlegm	3,076	−1.3	(−5.5, 2.9)	−1.6	(−5.6, 2.4)
	Wheezing attacks	3,076	0.7	(−2.3, 3.8)	1.0	(−1.9, 3.9)
	Trouble breathing	3,076	−4.4	(−9.9, 1.0)	−3.4	(−9.2, 2.4)
	Chest trouble	3,076	−4.2	(−8.8, 0.5)	−3.4	(−8.1, 1.3)
Measured (ppb)[c]	FeNO	793	0.2	(−2.3, 2.6)	0.3	(−2.2, 2.8)
Inflammatory Markers and MDA[d]						
Measured	IL-6 (pg/mL)	1,603	0.9	(−0.2, 1.9)	0.8	(−0.3, 2.0)
	TNF-α (pg/mL)	1,603	1.0	(0.1, 1.8)	0.8	(−0.1, 1.7)
	CRP (mg/L)	1,603	0.1	(−0.4, 0.6)	0.1	(−0.5, 0.6)
	MDA (μM)	1,603	0.3	(−0.1, 0.8)	0.2	(−0.2, 0.6)

ATT = average treatment effect on the treated; BP = blood pressure; CRP = C-reactive protein; DiD = difference-in-differences; ETWFE = extended two-way fixed effects; FENO = fractional concentration of exhaled nitric oxide; IL-6 = interleukin-6; MDA = malondialdehyde; Obs = observations; pp = percentage points; ppb = parts per billion; TNF-α = tumor necrosis factor-alpha.

[a] ETWFE models for BP models adjusted for age, sex, waist circumference, smoking, alcohol consumption, and use of BP medication.

[b] Self-reported respiratory outcomes adjusted for age, sex, smoking, occupation, frequency of drinking, and frequency of farming.

[c] Measured respiratory outcome (FeNO) adjusted for age, sex, body mass index, frequency of drinking, smoking category, and frequency of exercise, occupation, and time of measurement.

[d] Inflammatory marker and MDA outcome models adjusted for age, waist circumference, occupation, wealth index quantile, frequency of drinking, smoking category, and frequency of farming.

Mediated Impact on Health Outcomes

As noted earlier, we aimed to assess whether any health impacts of the CHP may work specifically through pathways involving changes in $PM_{2.5}$ and indoor temperature. Later, we show results from several mediation models. We focus the mediation analysis on the BP and respiratory outcomes for which we observed stronger evidence of total effects of the policy. We evaluated potential mediation for each mediator (indoor temperature and exposure to indoor $PM_{2.5}$) separately and in a single model accounting for multiple mediators, and we set the values of both mediators to the mean value for untreated participants at baseline (wave 1).

In **Table 9**, we show estimates of the *CDE*s for different sets of potential mediators. The first column of the first panel shows the covariate-adjusted total *ATT* of the CHP on brachial SBP (i.e., a total effect of a 1.4 mm Hg decrease as seen in

Table 9. Controlled Direct Effects for the Clean Heating Policy on Blood Pressure[a]

	Adjusted Total Effect		CDE Mediated by					
			Indoor $PM_{2.5}$		Indoor Temperature		Indoor $PM_{2.5}$ and Indoor Temperature	
	***ATT*[b]**	**(95% CI)**	***ATT*[c]**	**(95% CI)**	***ATT*[c]**	**(95% CI)**	***ATT*[c]**	**(95% CI)**
Brachial SBP (mm Hg)	–1.4	(–3.3, 0.5)	–0.8	(–2.9, 1.3)	–0.3	(–2.2, 1.6)	0.3	(–1.9, 2.5)
Central SBP (mm Hg)	–1.4	(–3.3, 0.4)	–0.8	(–2.9, 1.3)	–0.4	(–2.2, 1.3)	0.2	(–1.9, 2.4)
Brachial DBP (mm Hg)	–1.6	(–2.9, –0.3)	–1.1	(–2.7, 0.5)	–1.1	(–2.3, 0.1)	–0.6	(–2.1, 0.9)
Central DBP (mm Hg)	–1.6	(–2.9, –0.3)	–1.1	(–2.7, 0.6)	–1.2	(–2.4, –0.0)	–0.7	(–2.2, 0.9)

ATT = average treatment effect on the treated; CDE = controlled direct effect; CI = confidence interval; DBP = diastolic blood pressure; PM = particulate matter; SBP = systolic blood pressure.
[a] Results combined across 30 multiply imputed datasets (average of 3,082 observations per dataset).
[b] Adjusted for age, sex, waist circumference, smoking category, alcohol consumption, and use of blood pressure medication.
[c] Mediators were set to the mean value for untreated participants at baseline.

Table 8). The second panel shows the *CDE*, that is, the effect of exposure (vs. no exposure) to the CHP on brachial SBP in a counterfactual population in which we intervene to fix the value of indoor $PM_{2.5}$ to the average value for untreated participants at baseline. The *CDE* is –0.8 mm Hg (95% CI: –2.9, 1.3), demonstrating that under the assumptions outlined in the section "Data Analysis: Mediation Analysis," roughly 40% of the total effect of the CHP is mediated by the impact of the policy on indoor $PM_{2.5}$. The third panel shows the estimated *CDE* when holding constant the value of indoor temperature (without simultaneous adjustment for $PM_{2.5}$) to that of the untreated participants. This *CDE* is even smaller (–0.3 mm Hg, 95% CI: –2.2, 1.6), suggesting a somewhat stronger role for indoor temperature in mediating the total effect of the CHP on brachial SBP. Finally, the last panel shows a similar *CDE* of 0.3 mm Hg (95% CI: –1.9, 2.5) when setting both indoor $PM_{2.5}$ and indoor temperature to their respective pretreatment means. Holding the values of $PM_{2.5}$ and indoor temperature at preintervention values effectively eliminates these pathways by which the CHP can affect BP, so the small value of the *CDE* adjusting for both mediators suggests that the CHP effect on BP would be effectively null were it not for its effect on the mediators. Overall, the results in Table 9 indicate that conditioning on indoor $PM_{2.5}$ and indoor temperature largely explains the entire total effect of the CHP on SBP because the *CDE* conditional on both mediators was reduced to 0.3 mm Hg for brachial SBP. The *CDE*s for brachial and central DBP were roughly half the value of the total effect. Appendix Table 10 shows heterogeneous treatment effects for the mediation models for SBP and DBP, and are generally consistent with the patterns of mediation for the overall *CDE*s.

Table 10 shows estimates from similar analyses for the *CDE* of the policy on respiratory outcomes. For respiratory outcomes, we focus on mediation by personal exposure to $PM_{2.5}$ and point temperature; therefore, these estimates are derived for the subset of individuals with measures of personal exposure. Thus, the total adjusted *ATT*s in Table 10 are not directly comparable with those in Table 8, which included all study participants. We estimate the *CDE*s holding the values of both mediators to the average levels for never-treated households at baseline. Overall, we find no evidence that any of the total effects we observed for self-reported respiratory outcomes in Table 8 were mediated by personal exposure to $PM_{2.5}$ or indoor temperature. Generally, the *CDE*s for all the self-reported respiratory outcomes are statistically indistinguishable from the total effects estimated without controlling for mediators.

AIM 2: SOURCE CONTRIBUTIONS

Source analysis for this study was conducted using data from all eligible outdoor and personal exposure $PM_{2.5}$ samples. Eligible samples were those for which both $PM_{2.5}$ mass and chemical components were quantified. Individual chemical species concentrations (means and 95% CIs) for outdoor and personal samples by study wave are provided in Appendix Tables 24 and 25, respectively. We evaluated factors contributing to community-outdoor and personal exposure $PM_{2.5}$ using the US Environmental Protection Agency's source apportionment model positive matrix factorization (PMF) 5.0, which has been widely used for air pollution analyses in China (Gao et al. 2018; Liu et al. 2017; Tao et al. 2017). Because an optimum PMF result depends on the appropriate number of input factors, sensitivity analysis using a range of factors (e.g., 3–7 based on a combination of the measured chemical species, field observations, and sources previously identified in our study region) was conducted to examine the impact of a different number of factors on the model results. Detailed information on the procedures of PMF analysis can be found elsewhere (Wang et al. 2016; Zíková et al. 2016). Briefly, the scree plot from our principal component analysis indicated that solutions of between 3 and 5 factors (± 1) would be most appropriate, further supporting our evaluation of three- to six-factor solutions from PMF. Because there was

Table 10. Controlled Direct Effects of the Clean Heating Policy on Self-Reported Respiratory Outcomes[a]

	Adjusted Total Effect		CDE Mediated by					
			Personal $PM_{2.5}$		Indoor Temperature		Personal $PM_{2.5}$ + Indoor Temperature	
	ATT (pp)[b]	(95% CI)	*ATT (pp)*[c]	(95% CI)	*ATT (pp)*[c]	(95% CI)	*ATT (pp)*[c]	(95% CI)
Any symptom	–13.0	(–20.5, –5.5)	–12.7	(–20.7, –4.6)	–15.1	(–23.3, –6.9)	–15.0	(–23.7, –6.3)
Coughing	–10.5	(–19.2, –1.8)	–11.9	(–20.7, –3.1)	–14.5	(–23.1, –5.8)	–14.7	(–22.9, –6.6)
Phlegm	–9.5	(–16.6, –2.4)	–8.8	(–16.2, –1.4)	–14.7	(–24.3, –5.1)	–13.3	(–21.3, –5.4)
Wheezing attacks	–4.2	(–11.1, 2.6)	–2.8	(–8.7, 3.1)	–10.2	(–22.0, 1.6)	–3.8	(–10.0, 2.4)
Trouble breathing	–9.5	(–19.3, 0.3)	–9.6	(–19.6, 0.4)	–13.8	(–26.3, –1.3)	–13.1	(–25.2, –1.0)
Chest trouble	–5.0	(–12.2, 2.1)	–4.9	(–11.2, 1.4)	–4.6	(–11.8, 2.5)	–3.0	(–8.9, 2.9)

ATT = average treatment effect on the treated; CDE = controlled direct effect; CI = confidence interval; PM = particulate matter; pp = percentage point

[a] Estimated for the subset of individuals with measured personal exposure (n = 1,270).

[b] Adjusted for age, sex, smoking category, occupation, frequency of drinking, and frequency of farming.

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

no indication that even moving from five to six factors would improve our solution, we did not further investigate seven factors (**Table 11**).

Source Analysis Using Positive Matrix Factorization

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

The meteorological data included hourly boundary layer height, 2-m temperature, 2-m dew point temperature, and 2-m horizontal wind speed components (u, v), which were obtained from the fifth generation of the European Center for Medium-Range Weather Forecasts global climate reanalysis dataset (ERA5; 0.25° × 0.25° resolution). Values of these meteorological variables were determined at the village level by identifying the four surrounding grid points with values available from the ERA5 reanalysis and then applying inverse distance weighted interpolation from those four grid points to the village. Percent relative humidity was calculated from the 2-m dew point temperature using the "weathermetrics" package (version 1.2.2) in R (Anderson et al. 2016). Total hourly wind speed and wind direction were calculated from the horizontal wind speed components. A commonly accepted way to turn the two horizontal velocity components supplied by ERA5 (u = east-west, v = north-south, both in m s^{-1}) into the scalar wind speed and the meteorological wind direction is to treat u-v as the x- and y-coordinates of a vector and then apply simple vector algebra plus a four-quadrant inverse-tangent. In practice, we did this hour-by-hour for each village after the inverse-distance interpolation step.

The model diagnostics for the 3- to 6-factor pooled PMF solutions, which combined outdoor and personal exposure samples, are shown in **Table 11**. Model fit was assessed using a ratio of our model fit (Q) divided by the expected fit (Qexp). Because the change in Q/Qexp decreased with more factors, the model may be fitting additional sources that do

Table 11. Positive Matrix Factorization Error Estimation Diagnostics for Pooled Positive Matrix Factorization Analysis, Combining Outdoor and Personal Exposure Samples

	Potential Factor Solution			
Diagnostic	**3**	**4**	**5**	**6**
Qexp	27,936	26,052	24,168	22,284
Qtrue	187,681	147,796	123,236	100,316
Qrobust	174,407	139,910	117,082	95,932.5
Qr/Qexp	6.24	5.37	4.84	4.3
Q/Qexp >6	wi-Ca, ns-S, ws-Na, ws-Ca, Al, Cl, Pb	ns-S, Na, Al, Cl, Pb, Nitrate	Nitrate, ws-Na, Al, Chloride	Nitrate, ws-Na, Al
DISP % dQ	<0.1%	<0.1%	<0.1%	<0.1%
DISP swaps	0	0	0	0
BS_mapping	Dust: 98.5%	Transported dust: 95% Dust: 96.5% Sulfur secondary: 97.5%, Mixed combustion: 96.5%	Transported dust: 86% Mixed combustion: 87% Dust: 86% Lead: 55%	Transported dust: 84% Mixed combustion: 87.5% Dust: 81.5% Lead: 72% Chloride: 61.5% Sulfur secondary: 98.5%

wi = water-insoluble species; ws = water-soluble species; Qexp = expected fit; Qtrue = raw goodness-of-fit statistic (Σ scaled residual²) calculated with all observations included; Qrobust = goodness-of-fit statistic recalculated after the algorithm automatically down-weights any data points whose uncertainty-scaled residual |e/σ| > 4; Qr/Qexp = ratio of the robust fit to the statistically expected Q (≈ number of valid observations – number of free parameters); Q/Qexp = model fit divided by expected fit; DISP = displacement; DISP % dQ = maximum percentage increase in Q that occurred while each element of the base-case source profiles was systematically "displaced" (perturbed) during the DISP error-estimation routine; BS_mapping = bootstrap mapping

not improve the overall fit. The largest change in Q/Qexp was from three to four sources (6.24 to 5.37), but the changes moving from four to five factors and five to six factors were similar, which suggests that four factors are sufficient and parsimonious to explain the variation in our data. We assessed the random error in our model by randomly sampling blocks of data, fitting new models with the blocks, and comparing how the source profiles compared with the original model (bootstrap mapping). The three- and four-factor solutions had high bootstrap mapping (all factors identified in >96.5% of bootstrap runs). The additional sources identified in the five-factor (lead) and six-factor (chloride) solutions had low bootstrap mapping (>72%), indicating that these solutions are less consistent than the three- and four-factor solutions. The possibility that multiple solutions could result in the same Q value was assessed using displacement. The displacement approach takes the original factor profiles and modifies (±) the values for each species to maintain a small change in Q, reruns the solution with the new species values, and then compares the profiles of the new model with the original. Any swaps indicate that small changes to the species values could result in factor profiles that are different from the original solution, suggesting that the original solution is unstable. None of the factors in any of the solutions discussed were swapped during displacement, which indicates that all of the potential solutions are stable. Based on the Q/Qexp, bootstrap mapping, and interpretability of the factors, the four-factor solution was selected as most appropriate for the data.

The source profiles for the four-factor solution, which combined outdoor and personal exposure samples, are presented in **Figure 7**. We sought to develop a defensible source analysis solution, which we found to have four factors, and then identify and name those sources, jointly informed by our field observations, knowledge of local sources, and relevant previous studies. The pooled PMF analysis led to a more robust factor solution than personal or outdoor samples alone because of the increased number of samples. This pooled analysis determined that the optimum solution was a four-factor model, which identified the major sources as dust, transported dust, secondary sulfur, a combustion mixture of coal and biomass burning, and possibly some tobacco smoking. The pooled approach was found to be stable, with high bootstrap mapping and no factor swaps during the displacement tests, indicating the reliability of the factor solutions.

As sensitivity analyses, we additionally conducted a disaggregated analysis (Appendix Figure 12) in which personal

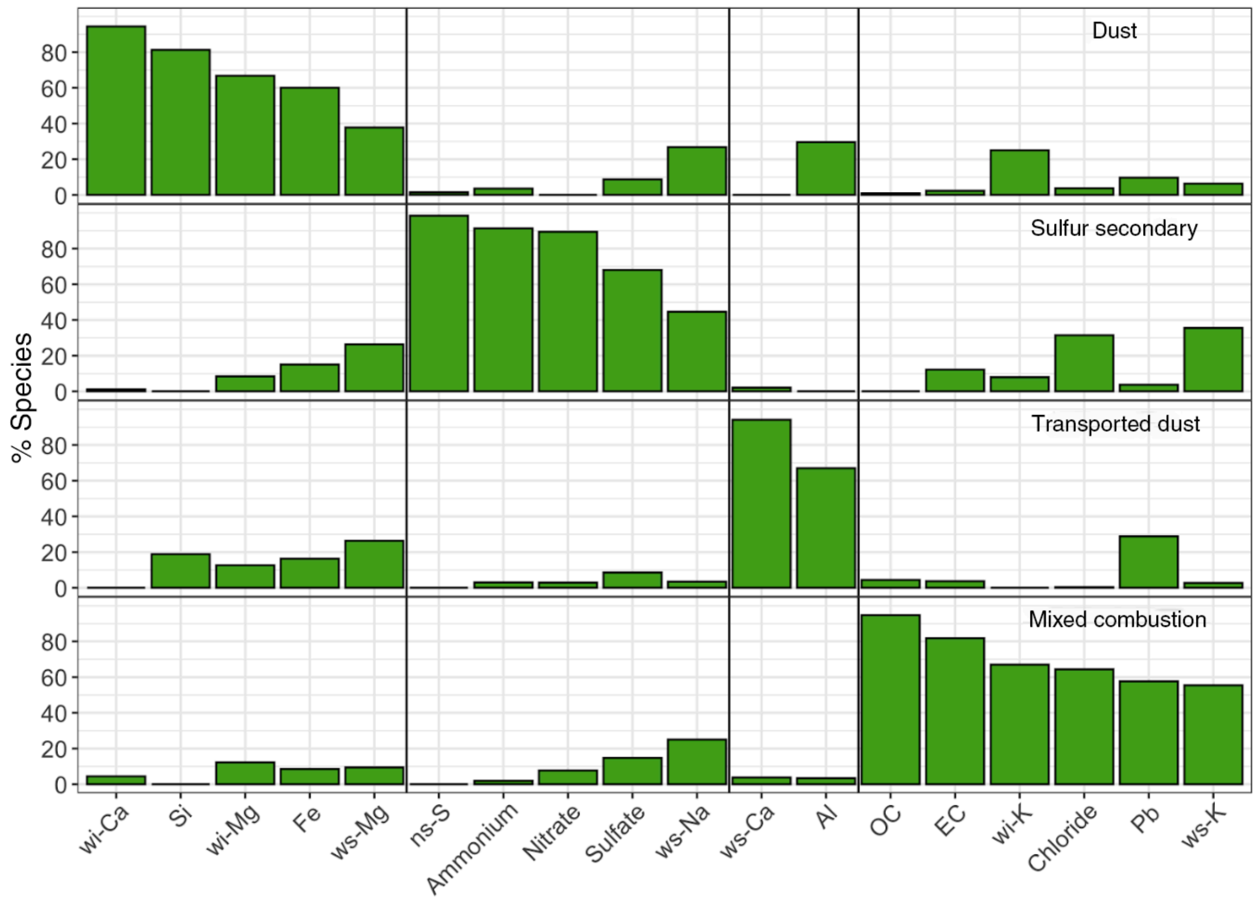


Figure 7. Source profiles for the four-factor positive matrix factorization solution to the sum of elements, ions, elemental carbon, and organic carbon for outdoor and personal $PM_{2.5}$ exposure measurements. Vertical lines separate the major contributing species to each source.

and outdoor samples were analyzed separately. We observed some differences in the number of factors and source attributions; however, the core findings remained largely consistent with the pooled results. For the personal exposure samples, whereas the three-factor solution identified dust, transported dust, and mixed combustion (a combination of biomass burning and secondary PM), the five-factor solution further split the mixed combustion factor into distinct coal combustion and sulfur secondary factors. Similarly, for outdoor samples, the three-factor solution identified mixed combustion, secondary PM, and dust, while the five-factor solution introduced transported dust and a refined characterization of secondary species contributions. The primary sources of pollution — dust, secondary sulfur, and mixed combustion — were consistent across pooled and disaggregated analyses. Thus, although the disaggregated analysis does provide additional granularity, it does not fundamentally change the identification of the key pollution sources that we observed in the pooled analysis.

We also evaluated PMF results disaggregated by day and by month (Appendix Figures 13 and 14), and the results are further color-coded by district. Because of yearly field campaign schedules, the timing of sampling in villages and districts was correlated. Therefore, this approach to source analysis does not yield results that allow us to disentangle changes in sources over time within seasons because they also potentially embed changes in sources across villages and districts in this study.

Thus, we concluded that the pooled analysis was the most parsimonious and interpretable approach for explaining the major sources of $PM_{2.5}$ in our study. The pooled results are both representative and stable, making them the most appropriate for addressing the study's primary research questions.

Description of $PM_{2.5}$ Sources Identified

The first source identified in the pooled PMF analysis was dust, characterized by high percentages of crustal elements such as wi-Ca, Si, and wi-Mg. The second source contained ns-S and secondary inorganic ions (ammonium, nitrate, and sulfate). Whereas ns-S is a tracer for primary coal combustion, secondary inorganic ions indicate a secondary source. Given the industrial coal burning in our study area, the secondary source likely combines primary and secondary emissions from coal and other sulfurous fuel combustion. Additionally, the higher outdoor concentrations of the secondary source compared with personal exposures support its identification as "sulfur secondary" because of its sunlight-driven secondary formation. The factor named "sulfur secondary" was intended to reflect the contribution of sulfur, as measured by the XRF analysis, that was not associated with sulfate. Had this contribution been coupled with other species associated with direct air pollutant emissions from coal and had those species not clustered with any other factors, we may have named this source a "household coal combustion" source. Finally, the fourth source included some species that are typical of direct air emissions from coal combustion and are also clustered with species that are typical of direct air emissions from biomass burning, leading to the naming of the factor as "mixed combustion." It is not surprising that residential coal and biomass emissions were difficult to fully separate because samples were not analyzed for organic tracers.

Factor 1, "dust," and Factor 3, "transported dust," represent different dust components. Northern China, including the Beijing region, regularly experiences large-scale dust storms that transport significant amounts of dust from distant desert regions such as the Gobi and Taklamakan Deserts, particularly in the spring. These dust storms, driven by strong winds from the northwest, often carry mineral dust high in silicon (Si) and iron (Fe) over thousands of kilometers. Satellite data and meteorological models consistently show that these dust storms result in the long-range transport of desert particles, which then settle over Beijing and surrounding areas, significantly impacting local air quality. For example, the 2021 dust storm (which overlapped with one of our field campaigns) was one of the strongest in recent memory, blanketing northern China with dust originating from the Taklamakan Desert, and highlighted the magnitude and distinct nature of these transported dust events compared with local dust sources (NASA Earth Observatory 2021).

Our ground-based observations and sensor network data corroborate the occurrence of such transported dust events. During these periods, we observed that elevated levels of PM could be traced to these long-distance sources, as indicated by chemical analysis. These dust storms bring particles that are chemically and physically distinct from those generated locally because local dust tends to have higher contributions from anthropogenic activities and nearer-surface sources of crustal materials.

Although both local and transported dust can contain elements such as Si and Fe, the transported dust's origin from vast desert regions gives it a unique composition and larger-scale influence. Studies of these storms have shown that transported dust often contains different proportions of these elements, shaped by the arid conditions and long-distance transport (Zhang et al. 2024a).

Based on our data and ground-truth experience, we believe that distinguishing transported dust from local dust is valid and reflective of the processes impacting air quality in the region. Moreover, previous studies have highlighted the importance of transboundary dust transport in shaping PM compositions in northern China, lending further support to our interpretation (Zhang et al. 2024b). The third source had high percentages of ws-Ca and Al, which in our study region is indicative of transported dust from dust storms that can occur in the spring. Although our samples were collected during the winter months only, it is possible that transported dust from previous years remained.

The fourth source was characterized by high percentages of tracers for both coal (OC, wi-K, chloride, Pb) and biomass combustion (EC, ws-K). Coal and biomass combustion are anticipated sources of $PM_{2.5}$ pollution in our study setting, particularly from domestic cooking and heating activities, so this source is likely a mixture of PM emitted from these two household combustion sources. We extend the source profiles across the different treatment cohorts in **Figure 8**.

Impact of Policy on Outdoor and Personal Exposure to the Mixed Combustion Source

Overall, **Table 12** shows that the average treatment effect of the CHP on outdoor (community) levels and personal exposure levels of the mixed combustion source was statistically indistinguishable from the null. Treatment was associated with lower, but statistically imprecise, personal exposures to the mixed combustion source. As with personal exposure to BC, an indicator of combustion pollution, this finding is consistent with the expectation that the policy contributed to reduced solid fuel emissions, as the "mixed combustion" source most likely reflects solid fuel combustion in our study setting. The results were consistent across the basic and covariate-adjusted models.

When the average treatment effects of the CHP policy on community outdoor levels and personal exposure levels of the mixed combustion source were allowed to vary by treatment year and time, the treatment effect for households most recently treated (i.e., treated in the final wave, wave 4) was associated with lower personal exposures to the mixed combustion source (Appendix Figure 11). In each wave, treatment by the CHP was associated with a reduction in the source contribution to personal $PM_{2.5}$ mass from the mixed combustion source; however, for villages treated in waves 2 and 3, the effect was statistically imprecise. Treatment was not associated with a reduction or an increase in the source contribution to community outdoor $PM_{2.5}$ mass from the mixed combustion source. Personal exposure measures of this specific air pollution source were found to be more indicative of treatment effect than community outdoor measures of the same source. This finding aligns with the expectation that the mixed pollution source, identified as a mixture of coal and biomass combustion, is characteristic of household use of solid fuels. These fuels, including coal and biomass, produce emissions that are likely to be closer to the people

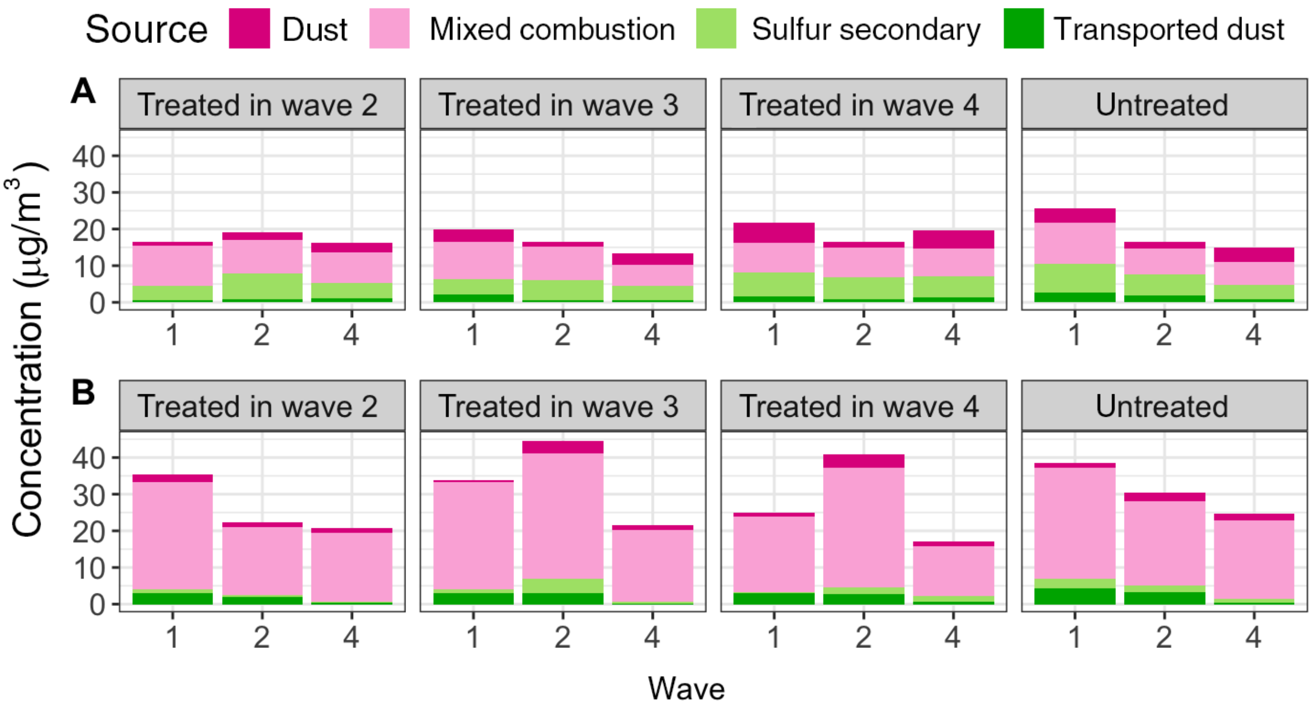


Figure 8. Arithmetic mean dispersion normalized source contributions found from the four-factor positive matrix factorization solution for (A) outdoor and (B) personal $PM_{2.5}$ exposure samples by year the group received treatment.

Table 12. Average Treatment Effect (μg/m³) for Outdoor and Personal Exposure to the Mixed Combustion Source

		DiD		Adjusted DiD	
	Obs	***ATT***	**(95% CI)**[a]	***ATT***	**(95% CI)**
Outdoor	717	1.07	(–4.90, 7.04)	1.53	(–4.19, 7.26)
Personal exposure	1,158	–5.60	(–13.70, 2.54)	–5.39	(–13.1, 2.35)

ATT = average treatment effect on the treated; CI = confidence interval; DiD = difference-in-differences; Obs = observations.
[a] Personal exposure model adjusted for temperature (represented by a spline with 2 degrees of freedom), participant smoking category, and whether the household reported using biomass fuel. Outdoor model adjusted for the total number of households in the village, total village population, and ambient relative humidity (represented by a spline with 2 degrees of freedom).

using them rather than near the centrally located community outdoor air samplers.

AIM 3: MEDIATION BY SOURCE CONTRIBUTION

Table 13 shows results from the mediation analysis by personal exposure to the mixed combustion source (coal and biomass), estimated for the subset of participants with personal exposure measurements. The *CDE* in this model estimates the impact of exposure to the CHP on central SBP and central DBP while holding constant values of mixed combustion source at the mean baseline values for the untreated population. The marginal policy effects (*ATTs*) from the adjusted ETWFE models for this subset of participants were the same as with the full sample for central SBP (–1.6 mm Hg) but slightly smaller for central DBP (–1.6 mm Hg in the full sample vs. –0.9 mm Hg in the subset with personal exposure measurements) and were estimated with greater imprecision in the smaller sample with personal exposure measurements. We found little evidence that these treatment effects were meaningfully mediated by exposure to the mixed combustion source because the controlled direct effects were generally of similar magnitude to the adjusted total effects.

DISCUSSION AND CONCLUSIONS

Air pollution emitted from residential space heating with coal has historically been a major contributor to cardiorespiratory disease burden in northern China (Archer-Nicholls

Table 13. Average Treatment Effects and Controlled Direct Effect (mm Hg) of the Clean Heating Policy on Central Systolic and Diastolic Blood Pressure with Mixed Combustion Source as the Potential Mediator

		DiD		Adjusted DiD		Adjusted CDE	
	Obs	***ATT***	**(95% CI)[a]**	***ATT***	**(95% CI)[b]**	***ATT***	**(95% CI)**
Central SBP	942	–1.6	(–5.3, 2.2)	–1.5	(–5.0, 2.0)	–1.5	(–5.0, 2.1)
Central DBP	942	–0.9	(–3.3, 1.5)	–1.1	(–3.2, 0.9)	–1.3	(–3.6, 0.9)

ATT = average treatment effect on the treated; CI = confidence interval; DiD = difference-in-differences; CDE = controlled direct effect; DBP = diastolic blood pressure; Obs = observations; SBP = systolic blood pressure.
[a] Adjusted for age, sex, waist circumference, smoking category, alcohol consumption, and use of blood pressure medication.
[b] Further adjusted for mediation by mixed combustion source (coal and biomass).

et al. 2016; Yun et al. 2020). Since the introduction of its 13th 5-Year-Plan (2016–2020), China has successfully implemented numerous large-scale measures to improve air quality, including programs that incentivize rural household transition from solid fuels to clean energy sources (Young et al. 2015). The CHP is among the largest and most ambitious household energy policies implemented anywhere in the world in recent decades, and its staggered rollout provided a unique opportunity to prospectively evaluate this real-world experiment and its effects on air quality and health.

ADOPTION OF THE HEAT PUMP TECHNOLOGY AND ADHERENCE TO THE POLICY

The CHP was successful in driving a rapid household heating energy transition from coal stoves to electric heat pumps in the treated study villages, with little difference in coal stove suspension or heat pump adoption for those treated before versus during the COVID-19 pandemic. There was high uptake and consistent use of the new heat pump technology, as well as large reductions in coal use in treated villages starting in the first year post-treatment and continuing into the third year of treatment for the villages first treated in 2019. We enrolled rural and peri-urban villages across a wide geographic area and socioeconomic spectrum in Beijing and observed near-universal adoption of the heat pump technologies and suspension of coal stove use across the different treatment groups and waves. This contrasts with many previous household energy intervention studies, including several randomized trials, in which low fidelity and compliance with the intervention stoves were considered major limitations to achieving their intended air quality or health benefits (Ezzati and Baumgartner 2017; Lai et al. 2024; Rosenthal et al. 2018).

Several factors contributed to the successful uptake of the new technology and adherence to the policy. The initial uptake of the heat pump technology was influenced by broad support and perceived benefits of village and household participation in the policy. At baseline assessment, 49 of 50 village committee interviewees indicated a desire to participate in the policy by the committee members and their constituents, for reasons including the ease of use of the heat pump, the convenience of no longer having to add coal throughout the day and especially the night, the desire for a cleaner local environment, and a perceived lower risk of carbon monoxide poisoning without coal stoves (data not reported). Although the availability and cost of clean fuels are well-established barriers to their adoption and sustained use over time (Rehfuess et al. 2014), in our study, both the upfront costs of the heat pump technology and a portion of electricity use were subsidized by the government, which limited the financial burden of clean energy transition for households. Furthermore, after policy implementation, treated villages no longer had access to government-subsidized coal, and household coal burning was further discouraged with possible punitive measures (e.g., potential loss of electricity subsidies).

IMPACTS OF THE POLICY ON HEALTH

One of the key findings from our comprehensive evaluation of the CHP was that exposure to the policy reduced SBP and DBP by approximately 1.5 mm Hg and that most of the observed BP effects were mediated by improvements in the indoor environment, specifically reductions in indoor $PM_{2.5}$ and increases in indoor temperature. The total effects of the policy are consistent with a small number of randomized trials of gas cookstoves or more efficient biomass cookstoves showing reductions in BP of similar magnitudes (Kumar et al. 2021). In contrast, recent randomized trials of LPG stoves in multiple countries observed no effect or a small (~0.6 mm Hg) increase in BP (Checkley et al. 2021; Ye et al. 2022) despite large decreases in personal exposures to $PM_{2.5}$ and BC. The inconsistency between our results and the LPG stove trial may stem from large age differences (mean ages 25 years and 48 years in the trials vs. 61 years in our sample) and the fact that gas stoves can still emit health-damaging air pollutants, including benzene and volatile organic compounds (Kashtan et al. 2023), especially in contrast with the zero-emission electric-powered heat pumps introduced to our study villages. Our findings of temperature- and air quality–mediated impacts of the policy on BP are also supported by observational studies showing that increased exposure to household air pollution (Baumgartner et al. 2018, 2011; Dong et al. 2013; Kanagasabai et al. 2022) and to colder indoor temperatures

(Lv et al. 2022; Sternbach et al. 2022) are associated with higher BP in rural and peri-urban areas of China, with exposure-response estimates that reasonably align with our estimates of the policy impact on BP after conditioning on temperature and $PM_{2.5}$ in the mediation analysis.

We did not observe effects of the policy on measures of PP or SBP amplification. PP is measured as the difference between SBP and DBP and represents the pulsatile component of blood flow (Dart and Kingwell 2001). Thus, increases in PP can result from increases in SBP, decreases in DBP, or both. The lack of effect on PP in our study is attributed to the similar reductions in SBP and DBP from the policy. Similarly, PP and SBP amplification are measured as a ratio of peripheral to central pressures, and the decreases in central and brachial pressures with the policy were also nearly identical in our study. Although the duration of our study was nearly twice as long as most previous household stove intervention studies conducted over 2 years or less, it is still possible that even longer-term reductions in BP are required to observe any structural changes in the caliber or elasticity of arterial walls that would subsequently be reflected in differences in PP or SBP and PP amplification (Dart and Kingwell 2001).

Our study also contributes to the limited evidence that a transition from solid fuel to clean energy can reduce the self-report of symptoms consistent with chronic respiratory tract irritation. Exposure to the CHP reduced self-report of any chronic respiratory symptoms by roughly 7 percentage points, with most of these effects driven by reductions in self-reported chest trouble or difficulty breathing on several or most days of the week. These findings align with previous randomized trials in Guatemala and Mexico, where biomass chimney stove interventions lowered indoor carbon monoxide and reduced the self-reported prevalence of chronic respiratory symptoms, especially coughing and wheezing, in younger women after 12 and 18 months of intervention (Romieu et al. 2009; Smith-Sivertsen et al. 2009). In contrast to our findings, the introduction of a solar cooker in Senegal provided no benefit to air pollution or self-reported respiratory symptoms (Beltramo and Levine 2013), and a recent trial of gas stoves in Peru similarly found no reduction in self-reported respiratory symptoms within the year of intervention despite very large reductions in $PM_{2.5}$ (Checkley et al. 2021).

We did not, however, find evidence that the reductions in chronic respiratory symptoms were mediated by changes in personal exposure to $PM_{2.5}$ or indoor temperature. This is not particularly unexpected because we did not observe an effect of the policy on personal exposure to $PM_{2.5}$, and any effects of temperature on chronic respiratory symptoms would more likely arise from large, rapid changes in temperature (D'Amato et al. 2018), but we observed small, gradual changes in our study. Future work will consider mediation by seasonal indoor $PM_{2.5}$, which is a longer-term measure of "usual" air pollution than 24-hour personal exposure and was shown to be reduced by the policy in our study homes.

We found some evidence of heterogeneity in the health benefits of the policy by treatment cohort. Generally, the policy showed strong reductions for the villages treated early and weak or null increases in BP for the last three villages exposed to the policy in 2021. We found less evidence for heterogeneity for self-reported respiratory symptoms, but did note potential evidence for increases in self-reported phlegm and wheezing attacks with the policy in the three villages treated in 2021. Notably, this was also the treatment cohort with the smallest improvement in point temperature at the time of BP measurement and an increase in self-reported biomass use. Paradoxically, we observed a larger decrease in $PM_{2.5}$ and mixed solid fuel use in this group. It is possible that the composition of PM and mixed solid fuel was different in this cohort, with a greater contribution of biomass smoke; however, we are unable to differentiate between biomass and coal in our "mixed combustion" category. Notably, this group of villages was also treated during the COVID-19 pandemic, which could have impacted how the policy was introduced in unpredictable and difficult-to-measure ways and could have also changed other health risk factors that we did not evaluate in our study (e.g., dietary intake).

We also found little evidence of an impact of the policy on blood biomarkers of inflammation and oxidative stress in the subsample of participants with blood collection in waves 1 and 2, but these were estimated with imprecision. Our results contrast with a natural experiment in urban Beijing that showed large regional and local air quality reductions during the 2008 Beijing Olympics and observed benefits to airway inflammation (Huang et al. 2012) and blood markers of inflammation and oxidative stress in healthy urban Beijing residents during the Olympics compared with before and after (Rich et al. 2012). Our mediation analysis indicated that the BP effects of the policy were mediated through both indoor temperature and air pollution, and the effects of the policy on SBP were mediated more through indoor temperature than air pollution. Although observational studies from rural northern China show impacts of exposure to temperature on inflammation and oxidative stress (Wang et al. 2020; Xu et al. 2019), it is possible that the relatively small increases in mean indoor temperature in treated households that we observed were not sufficiently large to capture measurable changes in these biomarkers.

IMPACTS OF THE POLICY ON AIR POLLUTION AND ITS SOURCES

The primary aim of the CHP was to reduce air pollution emissions and improve regional air quality, and it specifically targeted coal-burning stoves in northern China. Our evaluation of the CHP indicates a substantial improvement in indoor air quality, with a reduction of roughly –20.3 µg/m^3 (95% CI: –37.5,–3.0) in wintertime indoor $PM_{2.5}$ levels (Table 7). There is still considerable room for indoor air quality improvement, given that the seasonal indoor $PM_{2.5}$ levels in treated households in wave 4 (GM = 47 µg/m^3, 95% CI: 42, 52; Table 6) were still approximately 10 times higher than the annual WHO annual air quality guideline (5 µg/m^3) (WHO 2021).

Similar to our indoor results, several recent randomized trials with high compliance (exclusive or near exclusive) in the use of LPG stoves in rural settings with low outdoor pollution in Peru, Ghana, Guatemala, India, and Rwanda found lower exposures to $PM_{2.5}$ (32%–69%) in the intervention group compared with the control group using traditional solid fuel stoves (Checkley et al. 2021; Chillrud et al. 2021; Katz et al. 2020; Johnson et al. 2022), but even in these relatively low-pollution settings, postintervention mean exposures (range: 24–52 μg/m³) were still well above the WHO's annual air quality guideline (5 μg/m³). A trial in urban and peri-urban Nigeria, a high-pollution setting more similar to our Beijing sites, did not observe an air pollution benefit of ethanol stoves but did observe improved birth and pregnancy outcomes and BP (Alexander et al. 2017, 2018).

Nonetheless, comparisons of the indoor $PM_{2.5}$ benefits of the CHP in our study with previous assessments of household energy interventions in China suggest that the CHP performed well. In a separate study, homes with the NISP biomass chimney stoves had modestly lower indoor PM_4 than traditional open fire stoves (223 vs. 293 μg/m³), although postintervention air pollution levels were still an order of magnitude higher than the current health-motivated WHO (24-hour) guideline (Sinton et al. 2004). The NISPs' so-called "improved" coal heating stoves unexpectedly emitted higher concentrations of PM_4 and carbon monoxide than the traditional coal stoves (Edwards et al. 2004). In southwestern China (Sichuan), a DiD analysis of an government-supported household energy package pilot (semigasifier cookstove, water heater, pelletized biomass fuel) observed decreased indoor $PM_{2.5}$ (24%–67%) in women treated by the energy package, but greater reductions (48%–70%) were observed in untreated women, a result likely influenced by an unexpectedly large transition in gas cookstoves in untreated homes during the study period (Baumgartner et al. 2019).

The relatively high post-policy indoor air pollution levels and the limited benefit to personal exposures and outdoor $PM_{2.5}$ in treated villages in our study — despite excellent compliance with the policy — is likely due to three key factors. First, a quarter of our study households had at least one tobacco smoker, which is a large contributor to personal exposures to $PM_{2.5}$ in our study settings, especially during wintertime, when people tend to spend more time indoors (Li et al. 2022). Second, although Beijing has rapidly and impressively reduced outdoor $PM_{2.5}$ over the past decade (annual mean of 89 μg/m³ in 2013 decreased to 30 μg/m³ in 2022) (Zhang et al. 2023), the wintertime outdoor $PM_{2.5}$ levels in the treated study villages remained high enough across the study waves (range of means: 26–38 μg/m³ in treated villages) to limit the minimum exposure achievable with an indoor stove. The contribution of outdoor sources to personal exposures is further supported by our source apportionment analyses, which showed a clear contribution of regional sources (secondary sulfur, transported dust) to personal exposures. Finally, the continued use of biomass-burning kangs likely also contributed to indoor $PM_{2.5}$ and personal exposures. Kangs are a relatively simple and culturally entrenched combined cooking and space heating technique that has been used in China for more than 2,000 years (Zhuang et al. 2009). Kangs are mostly fueled by wood or other biomass that is freely and widely available in our study villages. The CHP did not ban biomass burning, and we observed persistent self-reported use of kangs after heat pump installation. Continued use of traditional solid fuel stoves alongside cleaner stoves and fuels (i.e., stove stacking) has long been a barrier to achieving large reductions in indoor and personal exposures after intervention (Shankar et al. 2020). A notable exception is the Household Air Pollution Intervention Network (HAPIN) trial, which attained near exclusive use of LPG stoves and dramatic reductions in personal exposures to $PM_{2.5}$ (lowered by 66% compared with control participants, 70 vs. 24 μg/m³) (Johnson et al. 2022), although the impressive air quality improvements were not accompanied by any health benefits across a range of neonatal, child, and maternal outcomes (Lai et al. 2024).

To comprehensively evaluate a large-scale policy such as the CHP, our study's measurement approach required extensive long-term measurements in more than 1,000 households in multiple waves using more than 500 air pollution monitors that collected thousands of hours of measurements. The scale and duration of air pollution measurement achieved in this study would not have been possible without low-cost air pollution sensors that have proliferated in the past decade. Our use of low-cost sensors to capture long-term (5–6 months) indoor air quality data in rural settings places it at the forefront of applying innovative technology to understand and mitigate household air pollution. This approach is somewhat unique for China because most studies, including those using lower-cost air pollution sensing networks (Chao et al. 2021; Mei et al. 2020), focus on urban air quality, driven by consideration for urban population demographic changes and industrial, power generation, and vehicular emissions (Shen et al. 2017). By focusing on rural indoor environments, our study addresses a crucial gap, offering insights into the effectiveness of a specific policy (CHP) on a microscale. Future evaluations of household energy interventions might also consider longitudinal measures of air pollution that track changes over longer periods to capture delayed effects. Estimating the causal effects of the CHP required a multifaceted approach to evaluation that incorporated a study design (DiD) and analytical methods (ETWFE, causal mediation) that are less common for evaluating air quality interventions. By incorporating a broader array of metrics and considering the systemic nature of air pollution and its health impacts, we sought to provide a more nuanced understanding of an intervention's effectiveness and how it may need to be augmented or restructured to achieve desired health outcomes.

ASSUMPTIONS, STRENGTHS, AND LIMITATIONS

The validity of our DiD approach is subject to two key assumptions (Callaway and Sant'Anna 2021; Wooldridge

2021). First, no anticipation: we assume that anticipation of the CHP did not affect outcomes before policy implementation and did not differ between treated and untreated villages. We selected villages that were eligible for the policy but not currently treated. It was generally understood that the policy would first be implemented in the plains areas with more updated electric grids and then gradually expand into more remote and mountainous areas of Beijing, although most of our study villages were far from Beijing's urban core. In addition to these geographical parameters, some of our study villages were assigned to the policy, whereas others applied to the local government, but they were generally unaware of if or when they would be treated at the time of enrollment. Second, parallel trends: our analysis assumes that in the absence of the policy, the trends in air quality and health in treated and untreated villages would have remained the same over time. Because the parallel trends assumption is based on a counterfactual, it cannot be empirically verified (similar to the assumption of no unmeasured confounding). However, given that we had two preintervention periods before the 2020 and 2021 cohorts were treated, we assessed the similarity of preintervention trends between waves 1 and 2 for the never-treated group and the cohorts eventually treated in 2020 and 2021. Estimates and 95% CIs for the difference in prepolicy trends are given in the Appendix for BP outcomes (Appendix Figure 15), personal exposure to $PM_{2.5}$ and BC (Appendix Figure 16), and self-reported respiratory outcomes (Appendix Figure 17). We did not find strong evidence for systematic differences in the prepolicy trends for any outcome, but some estimates were imprecise, and tests for prepolicy trends are generally considered to have low power (Roth 2022). In particular, we found some evidence that prepolicy trends in personal $PM_{2.5}$ were decreasing for never-treated villages but (imprecisely) increasing for the cohorts treated in 2020 and 2021. However, given that the (imprecise) cohort-specific *ATTs* for these two cohorts were in opposing directions (Appendix Table 6), diverging prepolicy trends seem an unlikely explanation for the (imprecisely) estimated heterogeneous treatment effects. Similarly, we found some evidence for differences in prepolicy trends for self-reported symptoms of coughing and chest trouble. In both cases, these symptoms appeared to be increasing between 2018 and 2019 somewhat faster for the cohort treated in 2021 relative to the never-treated group. This would lead to increases in the probability of these outcomes among the treated group, a pattern that is inconsistent with the negative cohort-specific *ATTs* reported in the Appendix. In addition, we adjusted for relevant time-varying confounders in estimating total effects and in mediation analyses, which aims to improve the credibility of the parallel trends assumption.

We also note that, in general, the addition of covariates to our "basic" ETWFE models did not meaningfully change our estimates. Nevertheless, we cannot entirely rule out the possibility that other programs or policies differentially affected air quality or health in treated and untreated villages, which could lead to over- or underestimation of their effects. To investigate this possibility, we surveyed a member of each village committee about other rural development or health policies and programs in their villages throughout our 4-year study period and did not identify any co-implemented programs that would differentially impact villages by treatment status and affect outcomes or mediators. Although some municipality- and district-level COVID-19–related preventive measures were implemented during the pandemic (e.g., lockdowns, travel restrictions), our study villages were not differentially exposed to any such measures. Finally, our mediation analysis assumes no residual time-varying confounding between our mediators (air pollution and temperature) and our health outcomes. Although we measured and evaluated a large number of time-varying risk factors for BP, we cannot eliminate the possibility of potential residual confounding, which could over- or underestimate the mediating effects of indoor environmental factors.

Strengths of this comprehensive, field-based assessment of the CHP include our quasi-experimental design to evaluate a clean energy intervention that would be near impossible to experimentally manipulate at the scale of our study. Our study design controlled for secular changes in health, and we additionally collected data on and adjusted for important time-varying covariates. It's perhaps worth noting that control for secular trends was important in our context. For personal exposure, supplementary analyses that dropped the time fixed effects or compared treated versus untreated villages with covariate adjustment would have suggested a substantial overestimate of the impact of the policy (Appendix Table 26). Our numerous sensitivity analyses showed the robustness of our findings to various analytic decisions. Most previous field-based household energy intervention studies were less than 2 years in duration with a single post-treatment wave (Lai et al. 2024; Quansah et al. 2017), and our 4-year study enabled longer-term evaluation of compliance with the coal ban and heat pump adoption and use, and their impacts on air pollution and health. Despite the logistical challenges of the COVID-19 pandemic shutdowns and related government restrictions that occurred throughout half of our study period, we were able to continue the study and successfully retain participants in all 50 study villages over 4 years. Our large sample size of 1,438 participants in 50 villages across multiple study waves enabled us to evaluate both the total effects of the policy and separately for different treatment cohorts. By comparison, the few previous field-based assessments of household energy interventions (trials and pre-post designs with controls) and BP ranged in size from 44 to 324 participants (Kumar et al. 2021; Lai et al. 2024; Onakomaiya et al. 2019), except the HAPIN trial, which enrolled approximately 3,000 pregnant women (Ye et al. 2022).

This study also has several limitations to consider when interpreting our results. First, the COVID-19 pandemic began in the middle of our study. Although there were no recorded COVID-19 infections in our study villages during the study period, Beijing's municipality-wide preventive measures impacted all study villages at the same time (e.g., travel restrictions, closure of public spaces, lockdowns, and quarantines). Roughly half of our treated villages entered into the policy during the pandemic, which likely had some influence on its rollout. We observed the largest benefits in BP and several respiratory outcomes in villages treated before the pandemic compared with those treated after it started. However, we cannot differentiate between treatment cohort effects attributable to treatment during the COVID-19 pandemic versus other factors that differed between treatment cohorts (i.e., geographic location, access to biomass, fuel prices).

Second, the CHP rollout began in 2016, but we did not begin enrolling villages into our study until 2018. Thus, many of our study villages are farther from the urban core and generally of lower socioeconomic status than many villages treated in the first 3 years of the policy. Previous studies of the CHP suggest that treated villages of all socioeconomic levels benefited from less-polluted and warmer indoor environments, but that the benefits were larger in wealthier villages that were more likely to use the heat pumps more often and set to a higher indoor temperature (Barrington-Leigh et al. 2019; Meng et al. 2023). Furthermore, most of our study villages had relatively easy access to (free) biomass fuel and may be more likely to use biomass-burning kangs than villages near the urban core, where biomass fuel is less readily available. Thus, our results may not be generalizable to all of rural and peri-urban Beijing, especially to the more urbanized, wealthier villages treated between 2016 and 2018, and may underestimate the impacts of the policy on indoor environmental factors that were important cardiorespiratory health mediators in our study.

Third, like any field-based study, we had a number of constraints with data collection. We were unable to measure indoor air quality in wave 1 because of logistical and budget constraints, and thus cannot directly estimate the effects of the policy on indoor $PM_{2.5}$ for the 10 villages treated in 2019, which is also the treatment cohort that experienced the greatest health benefits. Similarly, we were unable to collect blood samples in the last wave because all of our measurements were conducted in participants' homes rather than in clinics to avoid group contact during the pandemic. In addition, our study logistics required visiting 50 villages over a period of just several months. Thus, we were unable to return to villages if a previously enrolled participant was not at home at the time that staff visited the village. In such instances, we either randomly selected another eligible participant in the same home or randomly selected another household with eligible participants from the village roster, so our study participants differed slightly across waves. However, this is unlikely to impact our findings because our village-level study and analysis are designed to be robust to the participation of a random sample of participants in each wave, and there were few notable differences in key demographic characteristics or health behaviors between participants who contributed to a different number of waves or between participants across each of the three waves that included individual-level measurements.

Fourth, respiratory symptoms were self-reported, and thus, our estimated effects on respiratory symptoms must be interpreted with caution. Participants in our study were aware of their treatment status, and knowledge of being treated by a policy expected to reduce local air pollution may have affected their reporting of the perceived health benefits of intervention (Peel et al. 2015). Previous trials of improved biomass stoves, for example, noted a tendency of participants to report a favorable response to the stove regardless of its physiological efficacy (Burwen and Levine 2012; Smith-Sivertsen et al. 2009). Such reporting bias could have inaccurately increased the estimated effect of the policy, although we did not observe consistent effects across all respiratory outcomes or treatment cohorts, which one might expect if treated participants were inclined to give more favorable responses. Our study also benefited from the co-measurement of BP, an objective measure that is less likely to be biased by participant or staff awareness of treatment status, because our staff consistently followed strict quality control guidelines for measurement across all study villages and homes, regardless of treatment status.

Finally, some remarks concerning power and uncertainty. We do not distinguish our results based on traditional notions of "statistical significance" (Wasserstein et al. 2019), but given the large number of outcomes and wide range of reported estimates and uncertainties, questions about the power of our design are relevant. Using observed effect sizes to ask questions about power (sometimes called "post-hoc power calculations") is illogical given the direct relationship between power and observed *P* values or confidence limits (Hoenig and Heisey 2001). Instead, we aim to put our estimates in context using a retrospective design analysis (Gelman and Carlin 2014). Retrospective design analyses use plausible values for hypothetical effect sizes to ask about the strength of evidence a replicated study with our design and level of precision would be likely to provide. The main quantities for a design analysis include the probability that the replicated estimate would be "statistically significant" (power), the probability that the replicated estimate would have the wrong sign ("S-bias"), and the ratio of the replicated estimate divided by the true effect size (the "exaggeration ratio" or "M-bias"). We report these values for a range of conservative but plausible effect sizes.

For BP, we used a value of a 2.5-mm Hg decrease in SBP or a 2-mm Hg decrease in DBP (Baumgartner et al. 2011; Steenland et al. 2018). For self-reported respiratory outcomes, we assumed hypothetical effects on the order of a 10% decline for each outcome (translated into absolute percentage point declines), and for inflammatory markers, we hypothesized true effects of a roughly 5% decline (Pope et al. 2004; Tang et al. 2020), again translated into absolute terms.

The full results for all health outcomes are shown in Appendix Table 27. As an example, consider our estimate of the 7.5 percentage point decrease in any respiratory symptom (95% CI: 2.3, 12.7) (Table 8). Assuming that the true effect of the policy was a reduction of 5 percentage points, a replicated estimate with our design features and precision would have 47% power, virtually no chance of reporting the wrong sign (S-bias = 0), and an expectation that the estimated effect would be only 1.4 times too high. Similarly, if the true effect of the policy on brachial SBP was a reduction of 2.5 mm Hg, a replication of our study design would have roughly 73% power, a 2% chance of having the wrong sign, and an average exaggeration ratio of just 0.5, meaning it would be unlikely to exaggerate the true effect.

The greater precision for BP and respiratory outcomes lends greater confidence to these estimated impacts. On the other hand, our estimates of the impact on inflammatory markers contain greater uncertainty (larger standard errors), largely because of logistical constraints that prevented data collection in the last wave. For example, our adjusted *ATT* for the effect of the CHP on IL-6 was an increase of 0.8 pg/mL, but our data are compatible with effects as small as a 0.3-pg/mL decrease and as large as a 2.0-pg/mL increase. Appendix Table 27 shows that if the true effect of the CHP on IL-6 is a modest 0.2-pg/mL decline, a replication using our design and level of precision will have roughly 6% power, a 39% chance of having the wrong sign, and the estimated effect will, on average, be nearly 2.5 times too high. Because we chose generally conservative values for hypothetical effects, design analyses with larger hypothetical true effects would tend to show greater power, lower likelihood of reporting the wrong sign, and less exaggerated effects.

Because of the inherent uncertainty surrounding hypothesized "true" effects, it often makes sense to conduct retrospective design analyses for a range of effect sizes. We show estimates for BP reductions of 0–4 mm Hg in Appendix Figure 18. The HEI-funded objectives of this study were designed based on prestudy power calculations for BP reductions of roughly 0.25 standard deviations (~4 mm Hg for older rural population with an average SD of 16 mm Hg). Appendix Figure 18 shows that if the true effect of the policy were similar to our prestudy estimates of a 0.25 SD difference in SBP, a study with our level of precision would have greater than 90% power, virtually no chance of reporting an estimate with the wrong sign, and an average exaggeration ratio of just 1.04. This suggests that our design was well powered to detect clinically meaningful effects (Rahimi et al. 2021) on BP.

IMPLICATIONS OF FINDINGS

In this comprehensive field-based assessment of the rural CHP in Beijing, we observed high fidelity and compliance with the policy in our study villages and households, where nearly all households in treated villages stopped using coal and shifted to electric-powered heaters. Exposure to the policy reduced BP and self-reported chronic respiratory symptoms, and these health benefits were mediated by reductions in indoor $PM_{2.5}$ and improvements in home temperature. We did not observe the same benefits of the policy on outdoor air quality or personal exposures, likely because the relatively high contribution of other regional and local air pollution sources to outdoor and personal exposures may have masked the benefits from a single source reduction. We also did not observe the benefits of the policy on different measures of inflammation and oxidative stress in the subsample of participants with biomarker assessment, although we observed respiratory symptoms and BP benefits of the policy in a sensitivity analysis limited to the same participants. Still, our overall findings indicate that this ambitious policy achieved its goals in dramatically reducing residential coal burning and improving indoor air quality and temperature, which provided modest benefits to health.

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

DATA AVAILABILITY STATEMENT

The data and code that support the findings of this study are openly available in the Open Science Foundation's repository at *http://doi.org/10.17605/OSF.IO/8TWDS*, and the Standard Operating Procedures for all measurements are openly available in the Open Science Foundation's repository at *https://doi.org/10.17605/OSF.IO/9HKPQ*.

ACKNOWLEDGMENTS

In addition to the HEI funding that supported this work, we acknowledge support from the Canadian Institutes for Health Research (CIHR #159477) and the Social Sciences and Humanities Research Council (SSHRC #430-2017-00998 and #435-2016-0531). HEI funding supported the addition of indoor temperature and indoor air quality measurements starting in wave 2 to wave 4 and fully supported the data collection campaigns in waves 3 and 4. None of these funders had any role in study design, data collection and analysis, the decision to publish, or preparation of this report.

We thank and acknowledge the more than 1,400 study participants and the more than 50 field staff members who assisted with data collection and laboratory analysis. This work would not have been possible without the dedicated efforts and contributions of investigators and trainees who contributed to the development of ideas, data collection, and results that are reported here and in publications resulting from this work. Here, we wish to acknowledge Koren Mann, Arijit Nandi, Robert Platt, Kennedy Hirst, Enkhuun Byambadorj, Kaibing Xue, and Martha Lee. We also acknowledge the efforts of project coordinators in Canada and China: Xinwei

Liu, Jing Shang, Xiaoxia Hu, Jian Ma, Leona Siaw, Neha Ahmed, and Laojie Li.

REFERENCES

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

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

Alexander D, Northcross A, Wilson N, Dutta A, Pandya R, Ibigbami T, et al. 2017. Randomized controlled ethanol cookstove intervention and blood pressure in pregnant Nigerian women. Am J Respir Crit Care Med 195:1629–1639; doi:*10.1164/rccm.201606-1177OC*.

American Thoracic Society and European Respiratory Society (ATS/ERS). 2005. ATS/ERS recommendations for standardized procedures for the online and offline measurement of exhaled lower respiratory nitric oxide and nasal nitric oxide, 2005. Am J Respir Crit Care Med 171:912–930; doi:*10.1164/rccm.200406-710ST*.

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

Anderson GB, Peng RD, Ferreri JM. 2016. Weather metrics: Functions to convert between weather metrics (R package). Available: *https://cran.r-project.org/web/packages/weathermetrics/index.html*.

Archer-Nicholls S, Carter E, Kumar R, Xiao Q, Liu Y, Frostad J, et al. 2016. The regional impacts of cooking and heating emissions on ambient air quality and disease burden in China. Environ Sci Technol 50:9416–9423; doi:*10.1021/acs.est.6b02533*.

Arel-Bundock V. 2024. Marginal effects: Predictions, comparisons, slopes, marginal means, and hypothesis tests. R package version. 0.13.0. *https://CRAN.R-project.org/package=marginaleffects*.

Baker AC, Larcker DF, Wang CCY. 2022. How much should we trust staggered difference-in-differences estimates? J Finance Econ 144:370–395; doi:*10.1016/j.jfineco.2022.01.004*.

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

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

Baumgartner J, Clark S, Carter E, Lai A, Zhang Y, Shan M, et al. 2019. Effectiveness of a household energy package in improving indoor air quality and reducing personal exposures in rural China. Environ Sci Technol 53:9306–9316; doi:*10.1021/acs.est.9b02061*.

Baumgartner J, Schauer JJ, Ezzati M, Lu L, Cheng C, Patz JA, et al. 2011. Indoor Air pollution and blood pressure in adult women living in rural China. Environ Health Perspect 119:1390–1395; doi:*10.1289/ehp.1003371*.

Beltramo T, Levine DI. 2013. The effect of solar ovens on fuel use, emissions and health: Results from a randomised controlled trial. J Dev Eff 5:178–207; doi:*10.1080/19439342.2013.775177*.

Brauer M, Casadei B, Harrington RA, Kovacs R, Sliwa K; WHF Air Pollution Expert Group. 2021. Taking a stand against air pollution — the impact on cardiovascular disease: A joint opinion from the World Heart Federation, American College of Cardiology, American Heart Association, and the European Society of Cardiology. Circulation 143; doi:*10.1161/CIRCULATIONAHA.120.052666*.

Burwen J, Levine DI. 2012. A rapid assessment randomized-controlled trial of improved cookstoves in rural Ghana. Energy Sustain Dev 16:328–338; doi:*10.1016/j.esd.2012.04.001*.

Callaway B. 2020. Difference-in-differences for policy evaluation. In: Handbook of Labor, Human Resources and Population Economics (Zimmermann KF, ed). Cham, Switzerland: Springer International Publishing; 1–61.

Callaway B, Sant'Anna PHC. 2021. Difference-in-differences with multiple time periods. J Econom 225:200–230; doi:*10.1016/j.jeconom.2020.12.001*.

Cameron AC, Miller DL. 2015. A practitioner's guide to cluster-robust inference. J Hum Resour 50:317–372; doi:*10.3368/jhr.50.2.317*.

Card D, Krueger AB. 1994. Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. Am Econ Rev 84:772–793.

Chao C-Y, Zhang H, Hammer M, Zhan Y, Kenney D, Martin RV, et al. 2021. Integrating fixed monitoring systems with low-cost sensors to create high-resolution air quality maps for the northern China plain region. ACS Earth Space Chem 5:3022–3035; doi:*10.1021/acsearthspacechem.1c00174*.

Checkley W, Williams KN, Kephart JL, Fandiño-Del-Rio M, Steenland NK, Gonzales GF, et al. 2021. Effects of a household air pollution intervention with liquefied petroleum gas on cardiopulmonary outcomes in Peru. A randomized controlled trial. Am J Respir Crit Care Med 203:1386–1397; doi:*10.1164/rccm.202006-2319OC*.

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

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

Costello BT, Schultz MG, Black JA, Sharman JE. 2015. Evaluation of a brachial cuff and suprasystolic waveform algorithm method to noninvasively derive central blood pressure. Am J Hypertens 28:480–486; doi:10.1093/ajh/hpu163.

D'Amato M, Molino A, Calabrese G, Cecchi L, Annesi-Maesano I, D'Amato G. 2018. The impact of cold on the respiratory tract and its consequences to respiratory health. Clin Transl Allergy 8:20; doi:10.1186/s13601-018-0208-9.

Dai Q, Liu B, Bi X, Wu J, Liang D, Zhang Y, et al. 2020. Dispersion normalized PMF provides insights into the significant changes in source contributions to $PM_{2.5}$ after the COVID-19 outbreak. Environ Sci Technol 54:9917–9927; doi:10.1021/acs.est.0c02776.

Danesh J, Kaptoge S, Mann AG, Sarwar N, Wood A, Angleman SB, et al. 2008. Long-term interleukin-6 levels and subsequent risk of coronary heart disease: Two new prospective studies and a systematic review. PLoS Med 5:e78; doi:10.1371/journal.pmed.0050078.

Dart AM, Kingwell BA. 2001. Pulse pressure: A review of mechanisms and clinical relevance. J Am Coll Cardiol 37:975–984; doi:10.1016/S0735-1097(01)01108-1.

Dispersed Coal Management Research Group, 北京大学能源研究院气候变化与能源转型项目 (Climate Change and Energy Transition Project, Institute of Energy, Peking University). 2023. 中国散煤综合治理研究报告 (China Dispersed Coal Governance Report). https://www.ditan.com/static/upload/file/20231106/1699241777630902.pdf [accessed 18 April 2025]

Dockery DW, Rich DQ, Goodman PG, Clancy L, Ohman-Strickland P, George P, et al. 2013. Effect of Air Pollution Control on Mortality and Hospital Admissions in Ireland. Research Report 176. Boston, MA: Health Effects Institute.

Dominici F, Greenstone M, Sunstein CR. 2014. Science and regulation. Particulate matter matters. Science 344:257–259; doi:10.1126/science.1247348.

Dong G-H, Qian Z(Min), Xaverius PK, Trevathan E, Maalouf S, Parker J, et al. 2013. Association between long-term air pollution and increased blood pressure and hypertension in China. Hypertension 61:578–584; doi:10.1161/HYPERTENSIONAHA.111.00003.

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

Edwards RD, Smith KR, Zhang J, Ma Y. 2004. Implications of changes in household stoves and fuel use in China. Energy Policy 32:395–411; doi:10.1016/S0301-4215(02)00309-9.

Emerging Risk Factors Collaboration. 2012. C-reactive protein, fibrinogen, and cardiovascular disease prediction. N Engl J Med 367:1310–1320; doi:10.1056/NEJMoa1107477.

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

US Food and Drug Administration. 2018. Guidance Document: Bioanalytical Method Validation Guidance for Industry. Available: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/bioanalytical-method-validation-guidance-industry [accessed 10 February 2022].

Gao J, Wang K, Wang Y, Liu S, Zhu C, Hao J, et al. 2018. Temporal-spatial characteristics and source apportionment of PM2.5 as well as its associated chemical species in the Beijing-Tianjin-Hebei region of China. Environ Pollut 233:714–724; doi:10.1016/j.envpol.2017.10.123.

GBD MAPS Working Group. 2016. Burden of Disease Attributable to Coal-Burning and Other Air Pollution Sources in China. Special Report 20. Boston, MA: Health Effects Institute.

Gelman A, Carlin J. 2014. Beyond power calculations: Assessing type S (Sign) and type M (magnitude) errors. Perspect Psychol Sci 9:641–651; doi:10.1177/1745691614551642.

Goin DE, Riddell CA. 2023. Comparing Two-way fixed effects and new estimators for difference-in-differences: A simulation study and empirical example. Epidemiology 34:535; doi:10.1097/EDE.0000000000001611.

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

Gould CF, Bejarano ML, Kioumourtzoglou M-A, Lee AG, Pillarisetti A, Schlesinger SB, et al. 2023. Widespread clean cooking fuel scale-up and under-5 lower respiratory infection mortality: An ecological analysis in Ecuador, 1990–2019. Environ Health Perspect 131:037017; doi:10.1289/EHP11016.

Hoenig JM, Heisey DM. 2001. The abuse of power: The pervasive fallacy of power calculations for data analysis. Am Stat 55:19–24; doi:10.1198/000313001300339897.

Huang W, Wang G, Lu S-E, Kipen H, Wang Y, Hu M, et al. 2012. Inflammatory and oxidative stress responses of healthy young adults to changes in air quality during the Beijing Olympics. Am J Respir Crit Care Med 186:1150–1159; doi:10.1164/rccm.201205-0850OC.

Johnson M, Pillarisetti A, Piedrahita R, Balakrishnan K, Peel JL, Steenland K, et al. 2022. Exposure contrasts of pregnant women during the Household Air Pollution Intervention Network randomized controlled trial. Environ Health Perspect 130:097005; doi:10.1289/EHP10295.

Johnston FH, Hanigan IC, Henderson SB, Morgan GG. 2013. Evaluation of interventions to reduce air pollution from biomass smoke on mortality in Launceston, Australia: Retrospective analysis of daily mortality, 1994–2007. BMJ 346:e8446–e8446; doi:10.1136/bmj.e8446.

Kanagasabai T, Xie W, Yan L, Zhao L, Carter E, Guo D, et al. 2022. Household air pollution and blood pressure, vascular damage, and subclinical indicators of cardiovascular disease in older Chinese adults. Am J Hypertens 35:121–131; doi:10.1093/ajh/hpab141.

Kashtan YS, Nicholson M, Finnegan C, Ouyang Z, Lebel ED, Michanowicz DR, et al. 2023. Gas and propane combustion from stoves emits benzene and increases indoor air pollution. Environ Sci Technol 57:9653–9663; doi:10.1021/acs.est.2c09289.

Katz J, Tielsch JM, Khatry SK, Shrestha L, Breysse P, Zeger SL, et al. 2020. Impact of improved biomass and liquid petroleum gas stoves on birth outcomes in rural Nepal: Results of 2 randomized trials. Glob Health Sci Pract 8:372–382; doi:10.9745/GHSP-D-20-00011.

Keele L, Tingley D, Yamamoto T. 2015. Identifying mechanisms behind policy interventions via causal mediation analysis. J Policy Anal Manage 34:937–963; doi:10.1002/pam.21853.

Khuzestani RB, Schauer JJ, Wei Y, Zhang Y, Zhang Y. 2017. A non-destructive optical color space sensing system to quantify elemental and organic carbon in atmospheric particulate matter on Teflon and quartz filters. Atmos Environ 149:84–94; doi:*10.1016/j.atmosenv.2016.11.002*.

Kipen H, Rich D, Huang W, Zhu T, Wang G, Hu M, et al. 2010. Measurement of inflammation and oxidative stress following drastic changes in air pollution during the Beijing Olympics: A panel study approach. Ann NY Acad Sci 1203:160–167; doi:*10.1111/j.1749-6632.2010.05638.x*.

Kumar N, Phillip E, Cooper H, Davis M, Langevin J, Clifford M, et al. 2021. Do improved biomass cookstove interventions improve indoor air quality and blood pressure? A systematic review and meta-analysis. Environ Pollut 290:117997; doi:*10.1016/j.envpol.2021.117997*.

Lai AM, Clark S, Carter E, Shan M, Ni K, Yang X, et al. 2020. Impacts of stove/fuel use and outdoor air pollution on chemical composition of household particulate matter. Indoor Air 30:294–305; doi:*10.1111/ina.12636*.

Lai PS, Lam NL, Gallery B, Lee AG, Adair-Rohani H, Alexander D, et al. 2024. 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 209:909–927; doi:*10.1164/rccm.202402-0398ST*.

Lee M, Carter E, Yan L, Chan Q, Elliott P, Ezzati M, et al. 2021. Determinants of personal exposure to $PM_{2.5}$ and black carbon in Chinese adults: A repeated-measures study in villages using solid fuel energy. Environ Int 146:106297; doi:*10.1016/j.envint.2020.106297*.

Lewington S, LiMing L, Sherliker P, Yu G, Millwood I, Zheng B, et al. 2012. Seasonal variation in blood pressure and its relationship with outdoor temperature in 10 diverse regions of China: The China Kadoorie Biobank. J Hypertens 30:1383–1391; doi: *10.1097/HJH.0b013e32835465b5*.

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

Lindemann U, Stotz A, Beyer N, Oksa J, Skelton DA, Becker C, et al. 2017. Effect of indoor temperature on physical performance in older adults during days with normal temperature and heat waves. Int J Environ Res Public Health 14:186; doi:*10.3390/ijerph14020186*.

Liu B, Wu J, Zhang J, Wang L, Yang J, Liang D, et al. 2017. Characterization and source apportionment of $PM_{2.5}$ based on error estimation from EPA PMF 5.0 model at a medium city in China. Environ Pollut 222:10–22; doi:*10.1016/j.envpol.2017.01.005*.

Lowe A, Harrison W, El-Aklouk E, Ruygrok P, Al-Jumaily AM. 2009. Non-invasive model-based estimation of aortic pulse pressure using suprasystolic brachial pressure waveforms. J Biomech 42:2111–2115; doi:*10.1016/j.jbiomech.2009.05.029*.

Lv Y, Zhu R, Xie J, Yoshino H. 2022. Indoor environment and the blood pressure of elderly in the cold region of China. Indoor Built Environ 31:2482–2498; doi:*10.1177/1420326X221109510*.

Manning WG, Mullahy J. 2001. Estimating log models: To transform or not to transform? J Health Econ 20:461–494; doi:*10.1016/S0167-6296(01)00086-8*.

McCracken JP, Smith KR, Díaz A, Mittleman MA, Schwartz J. 2007. Chimney stove intervention to reduce long-term wood smoke exposure lowers blood pressure among Guatemalan women. Environ Health Perspect 115:996–1001; doi:*10.1289/ehp.9888*.

McCracken J, Smith KR, Stone P, Díaz A, Arana B, Schwartz J. 2011. Intervention to lower household wood smoke exposure in Guatemala reduces ST-segment depression on electrocardiograms. Environ Health Perspect 119:1562–1568; doi:*10.1289/ehp.1002834*.

Mei H, Han P, Wang Y, Zeng N, Liu D, Cai Q, et al. 2020. Field evaluation of low-cost particulate matter sensors in Beijing. Sensors 20:4381; doi:*10.3390/s20164381*.

Meng W, Zhu L, Liang Z, Xu H, Zhang W, Li J, et al. 2023. Significant but inequitable cost-effective benefits of a clean heating campaign in northern China. Environ Sci Technol 57:8467–8475; doi:*10.1021/acs.est.2c07492*.

Naimi AI, Kaufman JS, MacLehose RF. 2014. Mediation misgivings: Ambiguous clinical and public health interpretations of natural direct and indirect effects. Int J Epidemiol 43:1656–1661; doi:*10.1093/ije/dyu107*.

NASA Earth Observatory. 2021. Early Season Dust Storm Hits Beijing. *https://earthobservatory.nasa.gov/images/148052/early-season-dust-storm-hits-beijing*.

Ni K, Carter E, Schauer JJ, Ezzati M, Zhang Y, Niu H, et al. 2016. Seasonal variation in outdoor, indoor, and personal air pollution exposures of women using wood stoves in the Tibetan Plateau: Baseline assessment for an energy intervention study. Environ Int 94:449–457; doi:*10.1016/j.envint.2016.05.029*.

Niu J, Chen X, Sun S. 2024. China's coal ban policy: Clearing skies, challenging growth. J Environ Manage 349:119420; doi:*10.1016/j.jenvman.2023.119420*.

Olson MR, Graham E, Hamad S, Uchupalanun P, Ramanathan N, Schauer JJ. 2016. Quantification of elemental and organic carbon in atmospheric particulate matter using color space sensing — hue, saturation, and value (HSV) coordinates. Sci Total Environ 548–549:252–259; doi:*10.1016/j.scitotenv.2016.01.032*.

Onakomaiya D, Gyamfi J, Iwelunmor J, Opeyemi J, Oluwasanmi M, Obiezu-Umeh C, et al. 2019. Implementation of clean cookstove interventions and its effects on blood pressure in low-income and middle-income countries: Systematic review. BMJ Open 9:e026517; doi:*10.1136/bmjopen-2018-026517*.

Pearl J. 2000. Causality: Models, Reasoning, and Inference. Cambridge, UK: Cambridge University Press.

Pearson TA, Mensah GA, Alexander RW, Anderson JL, Cannon RO 3rd, Criqui M, et al. 2003. Markers of inflammation and cardiovascular disease: Application to clinical and public health practice: A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. Circulation 107:499–511; doi:*10.1161/01.cir.0000052939.59093.45*.

Peel JL, Baumgartner J, Wellenius GA, Clark ML, Smith KR. 2015. Are randomized trials necessary to advance epidemiologic research on household air pollution? Curr Epidemiol Rep 2:263–270; doi:*10.1007/s40471-015-0054-4*.

Pope CA III, Hansen ML, Long RW, Nielsen KR, Eatough NL, Wilson WE, et al. 2004. Ambient particulate air pollution, heart rate variability, and blood markers of inflammation in a panel of elderly subjects. Environ Health Perspect 112:339–345; doi:*10.1289/ehp.6588*.

Quansah R, Semple S, Ochieng CA, Juvekar S, Armah FA, Luginaah I, et al. 2017. Effectiveness of interventions to reduce household air pollution and/or improve health in homes using solid fuel in low-and-middle income countries: A systematic review and meta-analysis. Environ Int 103:73–90; doi:*10.1016/j.envint.2017.03.010*.

Rahimi K, Bidel Z, Nazarzadeh M, Copland E, Canoy D, Ramakrishnan R, et al. 2021. Pharmacological blood pressure lowering for primary and secondary prevention of cardiovascular disease across different levels of blood pressure: An individual participant-level data meta-analysis. Lancet 397:1625–1636; doi:*10.1016/S0140-6736(21)00590-0*.

Rehfuess EA, Puzzolo E, Stanistreet D, Pope D, Bruce NG. 2014. Enablers and barriers to large-scale uptake of improved solid fuel stoves: A systematic review. Environ Health Perspect 122:120–130; doi:*10.1289/ehp.1306639*.

Rich DQ, Kipen HM, Huang W, Wang G, Wang Y, Zhu P, et al. 2012. Association between changes in air pollution levels during the Beijing Olympics and biomarkers of inflammation and thrombosis in healthy young adults. JAMA 307; doi:*10.1001/jama.2012.3488*.

Ridker PM. 2001. High-sensitivity C-reactive protein: Potential adjunct for global risk assessment in the primary prevention of cardiovascular disease. Circulation 103:1813–1818; doi:*10.1161/01.cir.103.13.1813*.

Ridker PM, Hennekens CH, Buring JE, Rifai N. 2000. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. N Engl J Med 342:836–843; doi:*10.1056/NEJM200003233421202*.

Romieu I, Riojas-Rodríguez H, Marrón-Mares AT, Schilmann A, Perez-Padilla R, Masera O. 2009. Improved biomass stove intervention in rural Mexico: Impact on the respiratory health of women. Am J Respir Crit Care Med 180:649–656; doi:*10.1164/rccm.200810-1556OC*.

Rosenthal J, Quinn A, Grieshop AP, Pillarisetti A, Glass RI. 2018. Clean cooking and the SDGs: Integrated analytical approaches to guide energy interventions for health and environment goals. Energy for sustainable development. J Energy Inst 42:152–159; doi:*10.1016/j.esd.2017.11.003*.

Roth J. 2022. Pretest with caution: Event-study estimates after testing for parallel trends. Am Econ Rev 4:305–322; doi:*10.1257/aeri.20210236*.

RTI International. 2009. Standard Operating Procedure for the X-Ray Fluorescence Analysis of Particulate Matter Deposits on Teflon Filters: PM XRF Analysis. *https://www3.epa.gov/ttnamti1/files/ambient/pm25/spec/pmxrfsop.pdf*.

Rubin DB. 1987. Multiple Imputation for Nonresponse in Surveys. 1st ed. New York: Wiley. Available: *https://onlinelibrary.wiley.com/doi/book/10.1002/9780470316696*.

Rückerl R, Greven S, Ljungman P, Aalto P, Antoniades C, Bellander T, et al. 2007. Air pollution and inflammation (interleukin-6, C-reactive protein, fibrinogen) in myocardial infarction survivors. Environ Health Perspect 115:1072–1080; doi:*10.1289/ehp.10021*.

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

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

Secrest MH, Schauer JJ, Carter EM, Lai AM, Wang Y, Shan M, et al. 2016. The oxidative potential of $PM_{2.5}$ exposures from indoor and outdoor sources in rural China. Sci Total Environ 571:1477–1489; doi:*10.1016/j.scitotenv.2016.06.231*.

Shakya KM, Peltier RE. 2015. Non-sulfate sulfur in fine aerosols across the United States: Insight for organosulfate prevalence. Atmos Environ 100:159–166; doi:*10.1016/j.atmosenv.2014.10.058*.

Shang J, Zhang Y, Schauer JJ, Tian J, Hua J, Han T, et al. 2020. Associations between source-resolved $PM_{2.5}$ and airway inflammation at urban and rural locations in Beijing. Environ Int 139:105635; doi:*10.1016/j.envint.2020.105635*.

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

Shen H, Tao S, Chen Y, Ciais P, Güneralp B, Ru M, et al. 2017. Urbanization-induced population migration has reduced ambient $PM_{2.5}$ concentrations in China. Sci Adv 3:e1700300; doi:*10.1126/sciadv.1700300*.

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

Smith-Sivertsen T, Díaz E, Pope D, Lie RT, Díaz A, McCracken J, et al. 2009. Effect of reducing indoor air pollution on women's respiratory symptoms and lung function: The RESPIRE randomized trial, Guatemala. Am J Epidemiol 170:211–220; doi:*10.1093/aje/kwp100*.

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

Song C, Liu B, Cheng K, Cole MA, Dai Q, Elliott RJR, et al. 2023. Attribution of air quality benefits to clean winter heating policies in China: combining machine learning with causal inference. Environ Sci Technol 57:17707–17717; doi:*10.1021/acs.est.2c06800*.

Steenland K, Pillarisetti A, Kirby M, Peel J, Clark M, Checkley W, et al. 2018. Modeling the potential health benefits of lower household air pollution after a hypothetical liquified petroleum gas (LPG) cookstove intervention. Environ Int 111:71–79; doi:*10.1016/j.envint.2017.11.018*.

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

Sullivan AP, Holden AS, Patterson LA, McMeeking GR, Kreidenweis SM, Malm WC, et al. 2008. A method for smoke marker measurements and its potential application for determining the contribution of biomass burning from wildfires and prescribed fires to ambient $PM_{2.5}$ organic carbon. J Geophys Res Atmos 113; doi:*10.1029/2008JD010216*.

Sun L, Abraham S. 2021. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. J Econom 225:175–199; doi:10.1016/j.jeconom.2020.09.006.

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

Tang H, Cheng Z, Li N, Mao S, Ma R, He H, et al. 2020. The short- and long-term associations of particulate matter with inflammation and blood coagulation markers: A meta-analysis. Environ Pollut 267:115630; doi:10.1016/j.envpol.2020.115630.

Tao J, Zhang L, Cao J, Zhang R. 2017. A review of current knowledge concerning $PM_{2.5}$ chemical composition, aerosol optical properties and their relationships across China. Atmospheric Chem Phys 17:9485–9518; doi:10.5194/acp-17-9485-2017.

Thompson RJ, Li J, Weyant CL, Edwards R, Lan Q, Rothman N, et al. 2019. Field emission measurements of solid fuel stoves in Yunnan, China, demonstrate dominant causes of uncertainty in household emission inventories. Environ Sci Technol 53:3323–3330; doi:10.1021/acs.est.8b07040.

Tuck MK, Chan DW, Chia D, Godwin AK, Grizzle WE, Krueger KE, et al. 2009. Standard operating procedures for serum and plasma collection: Early Detection Research Network Consensus Statement Standard Operating Procedure Integration Working Group. J Proteome Res 8:113–117; doi:10.1021/pr800545q.

van Buuren S, Groothuis-Oudshoorn K. 2011. Mice: Multivariate imputation by chained equations in *R*. J Stat Softw 45; doi:10.18637/jss.v045.i03.

van Donkelaar A, Hammer MS, Bindle L, Brauer M, Brook JR, Garay MJ, et al. 2021. Monthly global estimates of fine particulate matter and their uncertainty. Environ Sci Technol 55:15287–15300; doi:10.1021/acs.est.1c05309.

VanderWeele TJ. 2015. Explanation in Causal Inference: Methods for Mediation and Interaction. New York: Oxford University Press.

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

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

Wang Y, Zhang Y, Schauer JJ, De Foy B, Guo B, Zhang Y. 2016. Relative impact of emissions controls and meteorology on air pollution mitigation associated with the Asia-Pacific Economic Cooperation (APEC) conference in Beijing, China. Sci Total Environ 571:1467–1476; doi:10.1016/j.scitotenv.2016.06.215.

Wasserstein RL, Schirm AL, Lazar NA. 2019. Moving to a world beyond "p <0.05." Am Stat 73:1–19; https://doi.org/10.1080/00031305.2019.1583913.

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

Wooldridge JM. 2021. Two-Way fixed effects, the two-way Mundlak regression, and difference-in-differences estimators. doi:10.2139/ssrn.3906345.

World Health Organization (WHO). 2021. WHO global air quality guidelines: Particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Geneva: World Health Organization. Available: https://www.who.int/publications/i/item/9789240034228/.

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

Xu W, Collet J-P, Shapiro S, Lin Y, Yang T, Wang C, et al. 2009. Validation and clinical interpretation of the St George's Respiratory Questionnaire among COPD patients, China. Int J Tuberculosis and Lung Disease 13:181–189.

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

Yang K. 2021. 电力工业为民族复兴提供不竭动力 (Power industry provides inexhaustible power for national rejuvenation.) 中国能源新闻网 (China Energy News Network).

Yap P-S, Garcia C. 2015. Effectiveness of residential wood-burning regulation on decreasing particulate matter levels and hospitalizations in the San Joaquin Valley air basin. Am J Public Health 105:772–778; doi:10.2105/AJPH.2014.302360.

Ye W, Steenland K, Quinn A, Liao J, Balakrishnan K, Rosa G, et al. 2022. Effects of a liquefied petroleum gas stove intervention on gestational blood pressure: Intention-to-treat and exposure-response findings from the HAPIN trial. Hypertension 79:1887–1898; doi:10.1161/HYPERTENSIONAHA.122.19362.

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

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

Yun X, Shen G, Shen H, Meng W, Chen Y, Xu H, et al. 2020. Residential solid fuel emissions contribute significantly to air pollution and associated health impacts in China. Sci Adv 6:eaba7621; doi:10.1126/sciadv.aba7621.

Zhang J, Smith KR. 2007. Household air pollution from coal and biomass fuels in China: Measurements, health impacts, and interventions. Environ Health Perspect 115:848–855; doi:10.1289/ehp.9479.

Zhang Q, Zheng Y, Tong D, Shao M, Wang S, Zhang Y, et al. 2019. Drivers of improved $PM_{2.5}$ air quality in China from 2013 to 2017. PNAS 116:24463–24469; doi:10.1073/pnas.1907956116.

Zhang S, Wu J, Yao J, Quan X, Zhai H, Lu Q, et al. 2024a. Dust monitoring and three-dimensional transport characteristics of dust aerosol in Beijing, Tianjin, and Hebei. Atmosphere 15:1212; doi:10.3390/atmos15101212.

Zhang W, Xu H, Yu X, Li J, Zhang Y, Dai R, et al. 2023. Rigorous regional air quality standards for substantial health benefits. Earth's Future 11:e2023EF003860; doi:*10.1029/2023EF003860*.

Zhang Z, Kuang Z, Yu C, Wu D, Shi Q, Zhang S, et al. 2024b. trans-boundary dust transport of dust storms in northern China: A study utilizing ground-based Lidar Network and CALIPSO Satellite. Remote Sens 16:1196; doi:*10.3390/rs16071196*.

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

Zigler CM, Kim C, Choirat C, Hansen JB, Wang Y, Hund L, et al. 2016. Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations. Research Report 187. Boston, MA: Health Effects Institute.

Zíková N, Wang Y, Yang F, Li X, Tian M, Hopke PK. 2016. On the source contribution to Beijing $PM_{2.5}$ concentrations. Atmos Environ 134:84–95; doi:*10.1016/j.atmosenv.2016.03.047*.

HEI QUALITY ASSURANCE STATEMENT

The conduct of this study was subjected to independent audits by RTI International staff members Dr. Linda Brown, Dr. David Wilson, and Mr. Ryan Chartier. These staff members are experienced in quality assurance (QA) oversight for air quality monitoring, modeling, and exposure assessment, epidemiological methods, and statistical modeling.

The QA oversight program consisted of a remote audit of the final report and the data processing steps. Key details of the dates of the audit and the reviews performed are listed below.

Audit: Final Remote Audit

Date: February 2025 – May 2025

Remarks: The final remote audit consisted of two parts: (1) review of the final project report and (2) audit of data processing steps. The review of the final report focused on ensuring that the methods are well documented and the report is easy to understand. The review also examined if the key study findings reported were supported by the data presented and if study limitations were discussed. The data audit included review of the datasets and codes for data reduction, processing and analysis, and comparison of the data outputs with reported data. This portion of the audit was restricted to the key components of the study and associated findings. Data and selected codes for exposure and epidemiological model development were sent to RTI.

The codes were reviewed at RTI to verify, to the extent feasible, linkages between the various scripts; confirmation of the models and model variables reported; and verification of key tables, figures, and data outputs. The codes and datasets appear to be largely consistent with the models described in the report and follow the overall model development procedure described. The values themselves were verified by RTI using the data and scripts provided by the investigators.

There were a few minor discrepancies, most of which were attributable to rounding differences or copy/paste errors; however, no major quality-related issues were identified from the review of the codes, data, and the report. Recommendations were made to address noted discrepancies and typographical errors, and included general edits for improved clarity. Those recommendations were addressed in the final report.

A written report was provided to HEI. The QA oversight audit demonstrated that the study was conducted according to the study protocol. The final report appears to be representative of the study conducted.

Linda Morris Brown, MPH, DrPH, Epidemiologist, Quality Assurance Auditor

David Wilson, PhD, Statistician, Quality Assurance Auditor

Ryan Chartier, MS, Air Quality and Exposure Scientist, Quality Assurance Auditor

Date: May 29, 2025

SUPPLEMENTARY APPENDIX ON THE HEI WEBSITE

Appendix A contains 27 tables and 18 figures not included in the main report. It is available on the HEI website at *www.healtheffects.org/publications*.

ABOUT THE AUTHORS

Jill Baumgartner, PhD, is a professor jointly appointed in the department of equity, ethics, and policy and the department of epidemiology, biostatistics & occupational health at McGill University. Her work evaluates the health impacts of air pollution, household energy use, and climate change.

Sam Harper, PhD, is a professor in the department of epidemiology, biostatistics & occupational health at McGill University. His research focuses on evaluating the impacts of social and economic policies on health.

Chris Barrington-Leigh, PhD, is an associate professor jointly appointed in the department of equity, ethics, and policy and the Bieler School of Environment at McGill University. His research investigates well-being, happiness, and sustainability, focusing on the socioeconomic and environmental factors that contribute to quality of life.

Collin Brehmer is a PhD student in the department of civil and environmental engineering at Colorado State University.

Ellison M. Carter, PhD, is an associate professor in the department of civil and environmental engineering at Colorado State University. Her research combines interests and expertise in air quality, exposure science, and chemistry, and aims to answer questions relevant to energy policy and its impacts on air pollution exposures and human health.

Xiaoying Li, PhD, is a research scientist in the department of mechanical engineering at Colorado State University and formerly a postdoctoral fellow at McGill University. Her research interests include investigating the interactions of indoor and outdoor air pollution and personal exposures to air pollution and evaluating the impacts of clean energy interventions on air quality and human health.

Brian E. Robinson, PhD, is an associate professor in the department of geography at McGill University who studies how people's livelihoods are influenced by ecosystem services and resource use, particularly in developing regions. His interdisciplinary research explores the interactions between livelihoods, the environment, and the institutions that govern resource management.

Guofeng Shen, PhD, is an assistant professor of environmental science at Peking University. His research interests and experiences are in sustainable household energy and environment, largely focusing on fates, impacts, and controls of hazardous pollutants produced from indoor solid fuel use, which is an important indicator of sustainable development.

Talia J. Sternbach is a PhD student in the department of epidemiology, biostatistics & occupational health at McGill University.

Shu Tao, PhD, is a professor of environmental science at Peking University who focuses on measuring and modeling clean energy transition, air pollution emissions, population exposures, and their estimated health impacts in China. His research integrates environmental science, atmospheric modeling, field measurements, and exposure assessment to better understand the sources, distribution, and health effects of air pollutants.

Kaibing Xue is a PhD student in atmospheric science at the University of the Chinese Academy of Sciences.

Wenlu Yuan is a PhD student in the department of epidemiology, biostatistics & occupational health at McGill University.

Xiang Zhang is a PhD student in the department of geography at McGill University.

Yuanxun Zhang, PhD, is a professor of atmospheric sciences at the University of the Chinese Academy of Sciences and directs the Yanshan Earth Critical Zone at the National Observation and Research Station in China. He has expertise in atmospheric chemistry and the measurement of air pollution and its chemical composition. His research focuses on understanding the chemical processes and mechanisms driving the formation of air pollutants and their impacts on climate and human health.

OTHER PUBLICATIONS RESULTING FROM THIS RESEARCH

Published

Li X, Baumgartner J, Barrington-Leigh C, Harper S, Robinson B, Shen G, et al. 2022. Socioeconomic and demographic associations with wintertime air pollution exposures at household, community, and district scales in rural Beijing, China. Environ Sci Technol 56:8308–8318; doi: *10.1021/acs.est.1c07402*.

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

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

Yuan W, Li X, Brehmer C, Sternbach T, Zhang X, Carter E, et al. 2025. Effects of outdoor and household air pollution on hand grip strength in a longitudinal study of predominantly rural Beijing adults. Int J Environ Res Public Health 22:1283, doi: *10.3390/ijerph22081283*.

In Preparation

Brehmer C, et al. Source contributions as an outcome for evaluating the impact of a household energy transition: An example from a coal-to-clean energy policy in Beijing, China.

Li X, et al. Performance of a low-cost sensor (LCS) network for long-term indoor and outdoor $PM_{2.5}$ measurement in rural communities in Beijing, China.

Gattie J, Yuan W, et al. Village committees and the local governance of China's Rural Clean Heating transition.

Gattie, J, et al. An evaluation of Beijing's Clean Heating Policy on adult health behaviours.

Sternbach TJ, et al. Impact of China's clean heating policy on indoor temperature.

Sternbach TJ, et al. Mediated impacts of China's clean heating policy on blood pressure by indoor air pollution and indoor temperature.

Yuan W, et al. Exposure-response relationships of outdoor and household air pollution with biomarkers of inflammation and oxidative stress in rural Chinese adults.

Zhang X, et al. Expenditure versus comfort: Distributional impacts on household space heating energy transition program in rural China.

Zhang X, et al. Household energy transition and subjective well-being: A difference-in-differences estimate of outcomes in rural northern China.

Xue K, et al. A difference-in-differences analysis evaluating the impact of a clean heating policy on inflammatory and oxidative stress biomarkers in Beijing adults.

Submitted or In Review

Sternbach T, Harper S, Li X, Brehmer C, Zhang Y, Yuan W, et al. China's clean heating policy and effects on blood pressure. Submitted to Nature Sustainability.

Brehmer C, Sternbach T, Byambadorj E, Hirst K, Barrington-Leigh C, Baumgartner J, et al. Impacts of a coal-to-clean heating policy on untargeted biomass stove use in rural Beijing. Submitted to Energy for Sustainable Development.

Xue K, Baumgartner J, Harper S, Sternbach T, Li X, Shang J, et al. Effects of China's Clean Heating Policy on respiratory symptoms and airway inflammation in adults: a quasi-experimental study in rural Beijing. In review at the American Journal of Respiratory and Critical Care Medicine.

Research Report 235, *How Do Household Energy Transitions Work?*, by J. Baumgartner and S. Harper et al.

INTRODUCTION

Protecting environmental quality and human health through air pollution actions or interventions typically incurs an economic cost. It is therefore important to understand whether environmental policies result in the intended improvements. This area of study, known as environmental accountability research (sometimes also referred to as intervention studies), evaluates the extent to which air quality actions or interventions have reduced air pollutant emissions and concentrations and improved public health. A major challenge in this research field is isolating changes that can be attributed to the actions in question from improvements that might be due to other actions or long-term trends. This challenge is of particular concern for actions that target numerous pollutant sources over large geographic regions and are implemented over several years.

Over the past two decades, HEI has played a prominent role as a leader in accountability research, contributing to research design, funding, study oversight, and evaluation of such research (see Preface). Through a series of Requests for Applications (RFAs),* HEI has now funded many studies that have assessed a wide variety of actions that have targeted both point and mobile sources of air pollution. Earlier studies tended to focus on local-level actions that were implemented over a relatively short time frame. HEI later solicited research that evaluated actions with a larger geographical scope or that were implemented over longer time frames.

In its 2018 research solicitation, *RFA 18-1* Assessing Improved Air Quality and Health from National, Regional, and Local Air Quality Actions, HEI aimed to fund empirical studies to assess the health effects of air quality actions or to develop methods required for, and specifically suited to, conducting such research and make them accessible and available to other researchers. Areas of interest included national- or regional-scale actions implemented over multiple years, local actions targeted at improving air quality in urban areas with well-documented air quality problems, and programs to improve air quality around major ports and transportation hubs and corridors.

In response, Baumgartner, Harper, and colleagues proposed to assess the effects of a household clean heating policy that mandated and subsidized 3,700 villages in the Beijing region of China to switch from highly polluting residential heaters fueled by coal to efficient electric- or gas-powered heat pumps between 2017 and 2021. They also proposed to assess potential behavioral, environmental, and chemical mechanisms that might explain how the policy affects health outcomes. The HEI Research Committee recommended the proposal for funding because it applies a strong study design to address an important policy and because it builds on an ongoing study.

This Commentary provides the HEI Review Committee's independent evaluation of the study. It is intended to aid the sponsors of HEI and the public by highlighting both the strengths and limitations of the study and by placing the results presented in the Investigators' Report into a broader scientific and regulatory context.

SCIENTIFIC AND REGULATORY BACKGROUND

EVALUATING THE EFFECTIVENESS OF HOUSEHOLD AIR POLLUTION INTERVENTIONS

It is well established that household air pollution from the combustion of solid fuels (e.g., coal and biomass) contributes to cardiorespiratory diseases and other adverse health effects (Chowdhury et al. 2023; HEI 2025; HEI Household Air Pollution Working Group 2018; Lai et al. 2024; WHO 2014). However, the extent to which household air pollution levels and health are improved by relevant interventions (e.g., replacing solid fuel stoves with more efficient electric or gas heaters) is unclear (HEI Household Air Pollution Working Group 2018; Lai et al. 2024). Given that as recently as 2022, about a quarter of the world's population (1.8 billion) relied on high-emission fuel sources for heating, there is a pressing need to understand how to effectively reduce exposures and improve health (IEA et al. 2025).

The effectiveness of household air pollution interventions is typically evaluated by intervention studies that compare measurements of air quality and health before and after an intervention, using randomized controlled trials or observational study designs. Most intervention studies of household

Dr. Jill Baumgartner's and Dr. Sam Harper's 3½-year study, "How Do Household Energy Transitions Work?" began in May 2020. Total expenditures were $1,094,118. The draft Investigators' Report was received for review in May 2024. A revised report, received in October 2024, was accepted for publication in November 2024. During the review process, the HEI Review Committee and the investigators had the opportunity to exchange comments and clarify issues in the Investigators' Report and its Commentary. Note: Review Committee members Frank Kelly, Jennifer Peel, and John Volckens did not partake in the review of the report due to conflicts of interest.

* A list of abbreviations and other terms appears at the end of this volume.

air pollution to date have focused on more efficient biomass cookstoves or changes to fuels, including liquified petroleum gas for cooking. These studies often measure fine particulate matter ($PM_{2.5}$; sometimes supplemented with measurements of black carbon or other major chemical constituents of $PM_{2.5}$) or carbon monoxide, and sometimes measure nitrogen dioxide (HEI Household Air Pollution Working Group 2018; Lai et al. 2024; Lee et al. 2020). The mixed results in the published literature about the effectiveness of these interventions have been attributed to study design choices, such as which health outcomes are assessed, the length of follow-up, and small sample sizes in some studies. Other issues have included that the adaptation and use of the interventions were context dependent and often limited (e.g., because the interventions were hard to use or broke), and that reductions in emissions from the intervention might be small relative to other sources of air pollution (HEI Household Air Pollution Working Group 2018; Lai et al. 2024).

Given the heterogeneous results of cookstove interventions and the more limited information on exposure to household air pollution from heating, studies are needed to provide credible and convincing estimates of the potential health benefits and the costs of interventions intended to accelerate transitions to clean household energy sources (HEI Household Air Pollution Working Group 2018). Such studies should address some of the limitations of the earlier cookstove studies by including other sources of household air pollution (e.g., home heating), evaluating the extent to which the new household energy technology has been adopted, and using prospective study designs to capture important pre-policy information (HEI Household Air Pollution Working Group 2018; Lai et al. 2024). Additionally, evidence on the mechanisms of how interventions produce changes in outcomes — a research question that can be assessed using a statistical approach known as mediation analysis — could be important to evaluate past interventions and design future interventions that are more effective (Keele et al. 2015).

HOUSEHOLD COAL COMBUSTION IN BEIJING

In suburban and rural parts of Beijing, fuel for heating during the cold winter season has historically consisted predominantly of coal and biomass. Households traditionally burned these fuels in standalone heating stoves and traditional kangs (i.e., raised platforms for both heating and cooking). Residential space heating in northern China required more than 200 million tons of coal in 2017 (Dispersed Coal Management Research Group 2023). The Global Burden of Disease — Major Air Pollution Sources (GBD MAPS) Working Group found that coal burning from residential, industrial, and power-generating sources contributed just over one-third of the total ambient wintertime $PM_{2.5}$ across China in 2013 (GBD MAPS Working Group 2016). According to GBD MAPS, emissions from residential coal combustion in Beijing alone were responsible for about 4.4 μg/m^3 of the annual average $PM_{2.5}$ concentrations in Beijing, accounting for 5.9% of the overall contributions to population-weighted $PM_{2.5}$ exposures and about 770 annual deaths related to air pollution in 2013. Separately, it has been estimated that residential coal combustion across the Beijing-Tianjin-Hebei region was responsible for about half of the wintertime $PM_{2.5}$ concentrations in Beijing before implementation of the clean heating policy, with lower contributions in other seasons (Zhang et al. 2013, 2017).

CHINA'S CLEAN HEATING POLICY

To reduce ambient air pollution concentrations, the Chinese government issued an *Air Pollution Prevention and Control Action Plan* in 2013 (China State Council 2013). Among other things, the Plan mandated (and subsidized) that over 1.5 million residents (up to 70% of all households) in rural areas of northern China, including the Beijing region, switch from highly polluting coal heaters to efficient electric or gas-powered heat pumps between 2017 and 2021. The government simultaneously designated coal-restricted areas and offered subsidies for electricity and electric-powered heaters, including electric heat pumps, to replace traditional coal-heating stoves. The clean heating policy included a pilot phase in 2015 and was rolled out on a village-by-village basis based on various factors related to policy priorities and local capacity starting in 2016.

EARLIER STUDIES EVALUATING CHINA'S CLEAN HEATING POLICY

Air quality and health improvements have been reported following the sweeping air pollution control policies that began in 2013, including the transition of cooking and heating fuels in the residential sector (e.g., Ding et al. 2019; Li et al. 2024; Zhang et al. 2019). The causal links between these air pollution regulations, emissions, ambient air pollution, and mortality have been assessed in a study led by Dr. Patrick Kinney with funding from HEI RFA 18-1 (Kinney et al. in press).

Because reducing the use of household coal was a major part of the air pollution control plan, researchers quickly initiated studies to evaluate the effectiveness of the clean heating policy in northern China. In a cross-sectional pilot study conducted in the winter of 2017, the investigators of the current study enrolled 302 households from three Beijing districts. Half of the households were located in three villages that had participated in the first wave of the clean heating policy, and the other half were located in three geographically contiguous villages in the same district that had not yet participated (Barrington-Leigh et al. 2019). They found that clean heating technologies completely replaced coal heaters in two of the villages with the policy, and partially replaced them in one of the villages with the policy. Indoor $PM_{2.5}$ concentrations were 67% lower and indoor temperatures were 1.4°C higher in villages participating in the policy compared with neighboring villages that were not participating. Additionally, in parallel with the current study, the investigators conducted a study that showed reductions in acute myocardial infarction in Beijing townships after roll-out of the clean heating policy (Lee et al. 2024). A separate study conducted by other investigators at the same time evaluated the clean heating policy in a multicity cohort of Chinese adults and found small decreases in the risk of chronic lung diseases (1.1% to 3.0%), but no change in cardiovascular disease risk after a one-year post-policy evaluation period (Wen et al. 2023).

MOTIVATION FOR THE CURRENT STUDY

The current HEI study builds on the ongoing Beijing Household Energy Transitions (BHET) study, which was intended to evaluate the effect of the clean heating policy on health. The observed differences in $PM_{2.5}$ concentrations and indoor wintertime temperatures in their earlier cross-sectional study motivated the investigators to delve deeper with difference-in-differences analyses that could account for stable characteristics that differ between the villages with and without the clean heating policy, as well as for shared trends in air pollution, health, and related factors over time. They therefore identified 50 suburban and rural Beijing villages that were eligible for the clean heating policy, but where the policy had not yet been implemented (**Commentary Figure 1**). The investigators expected about half the villages to implement the policy over the course of the BHET study. In December 2018, they enrolled 1,003 participants (about 20 per village) in 977 different households in the 50 participating villages. Trained staff collected information on socio-demographic characteristics, indoor temperature, outdoor and personal $PM_{2.5}$ exposures, and health starting in winter 2018–2019. They observed early indications of improved wintertime indoor $PM_{2.5}$ concentrations and indoor temperature after implementation of the policy.

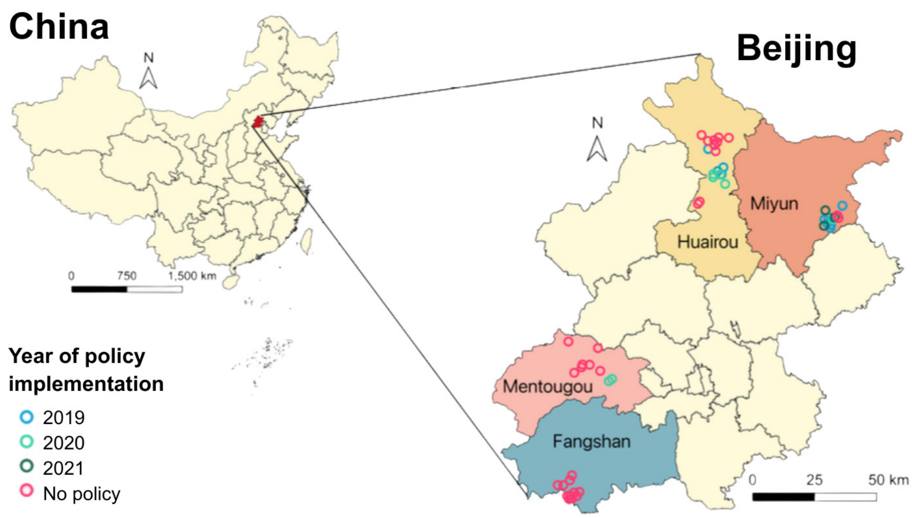
STUDY OBJECTIVES AND APPROACH

Baumgartner, Harper, and colleagues used HEI funding to extend the BHET study to assess the effects of the clean heating policy on outdoor and indoor air quality, personal exposures to air pollution, and cardiovascular and respiratory health. They also evaluated whether any changes in health outcomes observed after the implementation of the policy could be explained (i.e., were mediated) by changes in $PM_{2.5}$ concentrations and indoor temperature. Specific study aims were as follows:

1. Estimate the effect of the policy on outdoor, indoor, and personal $PM_{2.5}$ concentrations, and specifically estimate the source contribution from household coal burning to indoor and outdoor $PM_{2.5}$.
2. Estimate the overall effect of the policy on health, including respiratory symptoms and cardiovascular outcomes.
3. Assess the proportion of the observed effect of the policy on health that can be attributed to changes in $PM_{2.5}$, indoor temperature, and household coal burning.

To address these aims, the investigators measured many parameters related to air quality, temperature, and health in four consecutive winters from 2018 to 2022 (**Commentary Table 1**). They also collected information on household fuel use and self-reported respiratory health through participant questionnaires. To identify the contribution of household coal combustion to the $PM_{2.5}$ exposures, they conducted source apportionment of outdoor and personal measurements of $PM_{2.5}$ chemical composition.

To estimate the effects of the clean heating policy, the investigators conducted difference-in-differences analyses,



Commentary Figure 1. Map of village implementation of the clean heating policy. Each circle represents one recruited village. The colors of the circles indicate the year the clean heating policy was implemented in the villages. (Source: Adapted from Investigators' Report Figure 1.)

Commentary Table 1. Data Collection Across Waves in Four Consecutive Winters[a]

	Wave 1	Wave 2	Wave 3	Wave 4
Winter	2018–2019	2019–2020	2020–2021	2021–2022
Enrolled villages	50	50	50	50
Cumulative villages where the policy had been implemented	0	10	17	20
Household questionnaire	✓	✓		✓
Health measurements (blood pressure and cardiovascular biomarkers)	(✓)	(✓)		(✓)
Outdoor village-level $PM_{2.5}$	✓	✓	✓	✓
Indoor household $PM_{2.5}$		(✓)	(✓)	(✓)
Personal exposure to $PM_{2.5}$ and black carbon	(✓)	(✓)		(✓)
Chemical speciation of outdoor and personal $PM_{2.5}$	(✓)	(✓)		(✓)
Indoor temperature	✓	✓	✓	✓
Outdoor temperature	✓	✓	✓	✓
Confirmation of heating device use		(✓)	(✓)	(✓)

[a] ✓ indicates that the data were collected for the full population, and (✓) indicates that the data were collected for a subsample of the population as part of the study design. Numbers of measurements for each parameter and wave are included in the Investigators' Report. HEI funded Waves 3 and 4, and parts of Wave 2 (measurement of indoor $PM_{2.5}$, chemical speciation of outdoor and personal $PM_{2.5}$, and confirmation of heating device use using sensors).

a statistical approach used to compare changes over time in villages that had implemented the policy versus those that might implement it in the future. For health outcomes where they detected changes associated with the policy, the investigators then used mediation analysis to assess what proportion of the observed changes in health were caused by changes in air quality or temperature inside the home.

The investigators were forced to make some changes during the study because of the COVID-19 pandemic. Among the main changes were a partial campaign in winter 2020–2021 and the addition of a fourth full winter data collection campaign (including surveys, personal exposure, and health measurements) in winter 2021–2022.

SUMMARY OF METHODS AND STUDY DESIGN

STUDY AREA AND POPULATION

Local guides developed rosters of households that could be recruited into the study and recruited participants from about 20 randomly selected households in each of the 50 BHET villages for each study wave. About three-quarters of the recruited households remained in the study across all waves. Forty-three percent of individuals from the recruited households contributed observations across all three waves, and 31% of individuals participated in two waves. Individuals were eligible to participate if they were over 40 years old, lived in the study villages, were not planning to move out of the village in the next year, and were not currently receiving immunotherapy or corticosteroid treatments.

ENVIRONMENTAL EXPOSURE MEASUREMENTS

The study team performed detailed measurements for outdoor, indoor, and personal exposure to $PM_{2.5}$, along with outdoor and indoor temperatures. They measured $PM_{2.5}$ using both real-time sensors (one measurement per minute; Plantower) and gravimetric analysis of filters collected with Ultrasonic Personal Aerosol Samplers or Personal Exposure Monitors. All air pollution measurements followed established standard operating procedures with detailed quality assurance.

Community-level outdoor measurements of $PM_{2.5}$ were made using co-located sensors and filter samplers at one location near the center of each village and one or two locations at least 500 m away from the first measurement location. All outdoor measurement locations were distant from visible sources of $PM_{2.5}$. Co-located weekly filter samples were used to calibrate sensors and to measure the chemical composition of particles on the filters (i.e., elemental carbon, organic carbon, individual elements, and water-soluble ions).

Indoor measurements of $PM_{2.5}$ were made in six randomly selected households from each village starting with Wave 2. Sensors were deployed between late November and mid-January in the rooms where participants reported spending most of their time. The sensors made continuous measurements until they were collected in late April. Filter-based samplers were co-located with a subset of the indoor sensors during the first 24 hours of sensor-based measurements to aid in sensor calibration.

Personal exposures to $PM_{2.5}$ and black carbon were assessed for about 500 participants in each wave (10 randomly selected study participants in each village) who each carried a filter sampler for 24 hours. Outdoor temperature and relative humidity were obtained from the Beijing meteorological network. Additional meteorological data (e.g., boundary layer height) were obtained from the fifth generation of the European Center for Medium-Range Weather Forecasting Forecasts global climate reanalysis dataset (ERA5).

The investigators also measured indoor air temperature in a centrally located room in each house for the 5 minutes before blood pressure measurements and, for 75% of households, every 125 minutes for up to 4 months in the winter and early spring to check which heating and cooking devices were in use.

$PM_{2.5}$ SOURCE APPORTIONMENT

The investigators used a source apportionment approach to determine the contribution of household coal burning to the combined set of outdoor and personal $PM_{2.5}$ samples from which both $PM_{2.5}$ mass and chemical components were quantified. The investigators used the United States Environmental Protection Agency's source apportionment model, positive matrix factorization (PMF) 5.0, which has been widely used for air pollution analyses in China. They accounted for differences in wind speed and boundary layer height in the chemical analysis data prior to their use in the PMF model.

They conducted sensitivity analyses using source apportionment models with different measurement subsets and three to six source factors and evaluated the fit of the final PMF model using standard PMF model diagnostics. Factors identified from the PMF analyses were interpreted and named based on the field observations, investigators' knowledge of local sources, and relevant previous studies.

HEALTH OUTCOMES

The investigators focused on various cardiovascular outcomes and respiratory symptoms that have well-established links to $PM_{2.5}$. Blood pressure (brachial systolic, brachial diastolic, central systolic, and central diastolic) and chronic airway symptoms were assessed at the participants' homes at the time of the questionnaire visits (see below). Blood pressure was measured by trained staff using factory-calibrated devices with standard protocols and following detailed quality control procedures that included an appropriately sized cuff, correct positioning of the arm, and having both feet on the ground. Self-reported airway symptoms were assessed using a questionnaire (details below) with standard validated questions about chronic airway symptoms, including cough, phlegm, wheeze, and tightness in the chest. Additionally, the study team measured airway inflammation (as the fractional concentration of exhaled nitric oxide [FeNO]) in about one-quarter of participants.

Within 1 month of the researchers' visit, participants visited a village clinic where the study team collected fasting blood samples. Samples were tested for glucose, a complete lipid profile, and biomarkers of systemic inflammation or oxidative stress (C-reactive protein, interleukin-6, tumor necrosis factor alpha, and malondialdehyde) that are associated with both exposure to air pollution and the development of cardiovascular disease and events. Body weight, height, and waist circumference were also measured.

QUESTIONNAIRES

In addition to the respiratory symptoms described above, the study team collected information on households and individual participants by administering structured questionnaires in Mandarin Chinese. They gathered information on the types of fuels and stoves present, patterns of fuel and stove use, and the amount of fuel used for space heating in winter. They also collected supporting information on socio-demographic factors, house structure, and individual lifestyle factors (e.g., smoking). One representative from each village completed a survey about the policy implementation and the level of village interest in the policy, as well as information about any other policies being implemented simultaneously.

STATISTICAL ANALYSES

Measuring Effects of the Clean Heating Policy

To estimate the effects of the policy, the investigators conducted difference-in-differences analyses using multivariable "extended" two-way fixed effects models to account for the staggered roll-out of the policy, as described by Wooldridge (2021). This design compared changes in outcomes among villages that had the policy implemented in the same year, different years, or not at all. The villages where the policy was not implemented by the end of the study period provided an estimate of what changes in the health and exposure outcomes would have been expected to happen in the other villages if they had not implemented the policy (a "counterfactual"). By comparing changes in outcomes among villages with and without the policy, the difference-in-differences approach accounts for both time-invariant characteristics of each village as well as trends in air quality and health over time that are unrelated to the policy. As a result, these analyses produce estimates of the total effect of the policy on air quality and health (**Commentary Figure 2, top**).

Comparing Direct and Indirect Effects of the Clean Heating Policy

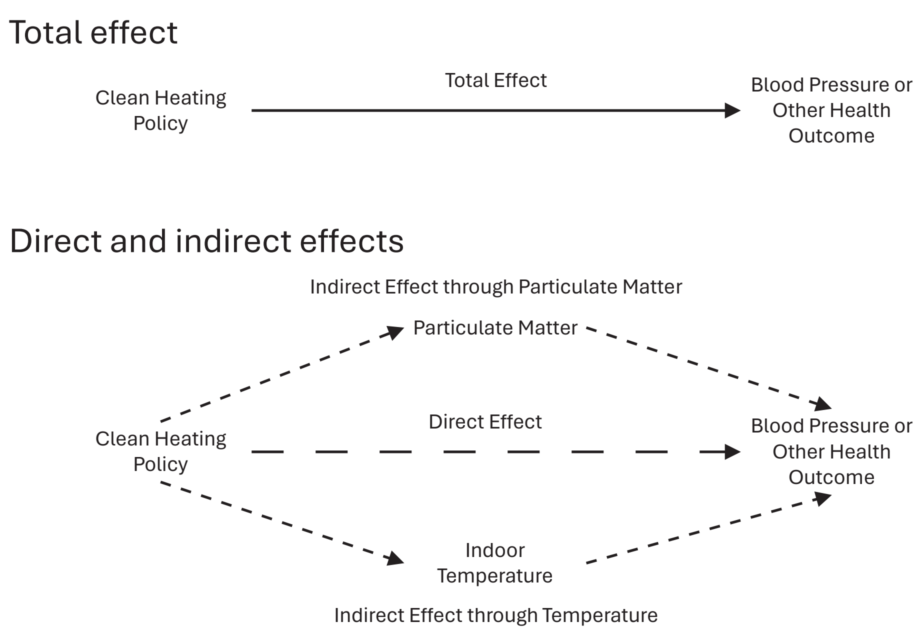
An important part of the study was to evaluate whether the clean heating policy influenced health directly, or indirectly through its effects on other factors (known as mediators), which then affect health. To do this, the investigators applied an approach for mediation analysis in the causal inference framework (Hernán and Robins 2020; Keele et al. 2015).

For those outcomes where they found a total effect, the investigators estimated the contributions of potential mediators of those associations (**Commentary Figure 2, bottom**). The potential mediators they considered were indoor or personal $PM_{2.5}$ concentrations and indoor temperature. The investigators tested the influence of these potential mediators by repeating the initial models, but with the potential mediator being tested held constant for the entire study population. This analysis produced the "direct effect" of the policy to describe the relationship between the policy and health outcomes independent of the potential mediators. Finding that the estimated direct effect was smaller than the total effect would suggest that the effect of the clean heating policy on health was at least partly because of its effect on the mediating factor being tested (in other words, that there was mediation). Conversely, finding that the direct and total effects were similar would suggest that the health benefits of the clean heating policy were related to factors other than changes in measured indoor $PM_{2.5}$ concentrations or measured temperatures, such as residents going outside less frequently to refuel the heaters or more even heating within the homes.

Adjustment for Potential Confounders and Imputation of Missing Data

The investigators identified potential confounders using directed acyclic graphs (DAGs) and included different sets of adjustments for potential confounding in the models for each outcome. To obtain valid results, they adjusted the total effect and direct effect models for confounding by factors such as participant age, sex, exposure to tobacco smoke, height, and weight. They also considered potential confounders related to the size of the villages and homes, and related to other factors that might vary over time and differ among villages.

Several key covariates for blood pressure models (e.g., waist circumference, height, and weight) were missing from 15% of participants in Waves 1 and 2 because those individuals did not participate in the clinic visits. Also, indoor $PM_{2.5}$ concentrations were measured at only about one-fifth of the homes in each wave by design, so 80% of the participants had no matching indoor $PM_{2.5}$ concentrations. There was little overlap between participants with blood pressure measurements and participants whose homes had been sampled for indoor $PM_{2.5}$. The investigators therefore used multiple



Commentary Figure 2. Simplified diagrams of the main models in the study. Top: The total effect of the policy, including all mechanisms by which the policy might work. Bottom: Separating indirect effects through changes in two potential mediators (particulate matter and indoor temperature; dotted lines) from the direct effect of the policy (dashed line), which includes all other possible mechanisms.

imputation with chained equations to replace the missing covariates and indoor $PM_{2.5}$ concentrations so that they could make full use of the blood pressure data.

SUMMARY OF KEY RESULTS

STUDY POPULATION

The investigators enrolled 1,438 participants over the study period (**Box 1: A Comprehensive Study**). Demographic characteristics, health, $PM_{2.5}$ concentrations, and indoor temperature were similar at baseline for participants in villages where the policy was implemented by the end of the study and where it was not. The average age of study participants at baseline (i.e., in Wave 1) was 60 years. About 60% of participants were female and about 80% were current or former smokers or lived with someone who smoked. On average, the participants had slightly elevated blood pressure (130 mm Hg systolic blood pressure and 82 mm Hg diastolic blood pressure), and just over half of the participants reported at least one respiratory problem, usually shortness of breath or phlegm. There was little missing data, and the investigators did not find a relationship between the implementation of the policy and missing data.

Box 1: A Comprehensive Study

Large Sample Size

- 1,438 participants from 1,236 households in 50 villages visited over four winters

Detailed Characterization of Environmental Exposures

- 11,174 outdoor, 399 indoor, and 1,270 personal sensor-based 24-hour $PM_{2.5}$ measurements
- 968 outdoor, 288 indoor, and 1,295 personal $PM_{2.5}$ filter samples (a subset of which were also used to quantify chemical constituents of $PM_{2.5}$)
- 1,161 personal measurements of black carbon on filters
- 717 outdoor and 1,158 personal estimates of exposure to "mixed combustion" from source apportionment
- 2,999 measurements of 5-minute indoor temperature

Individual Health Outcomes

- 3,082 measurements of blood pressure
- 3,076 measurements of self-reported respiratory effects (coughing, phlegm, wheezing attacks, trouble breathing, and chest trouble)
- 793 measurements of exhaled nitric oxide (FeNO)
- 1,603 measurements of inflammatory markers (C-reactive protein [CRP], interleukin-6 [IL-6], tumor necrosis factor-alpha [TNF-α]), and an oxidative stress marker (malondialdehyde [MDA]) in blood

POLICY UPTAKE

During the study, villages where the policy was implemented had almost complete uptake and adherence to the policy, as shown by the shift to heat pumps as the primary heating device (**Commentary Figure 3, top**). About one-third of households from treated villages that previously used coal stoves replaced the coal stoves with exclusive heat pump usage, and the rest used a mixture of heat pumps with biomass kangs.

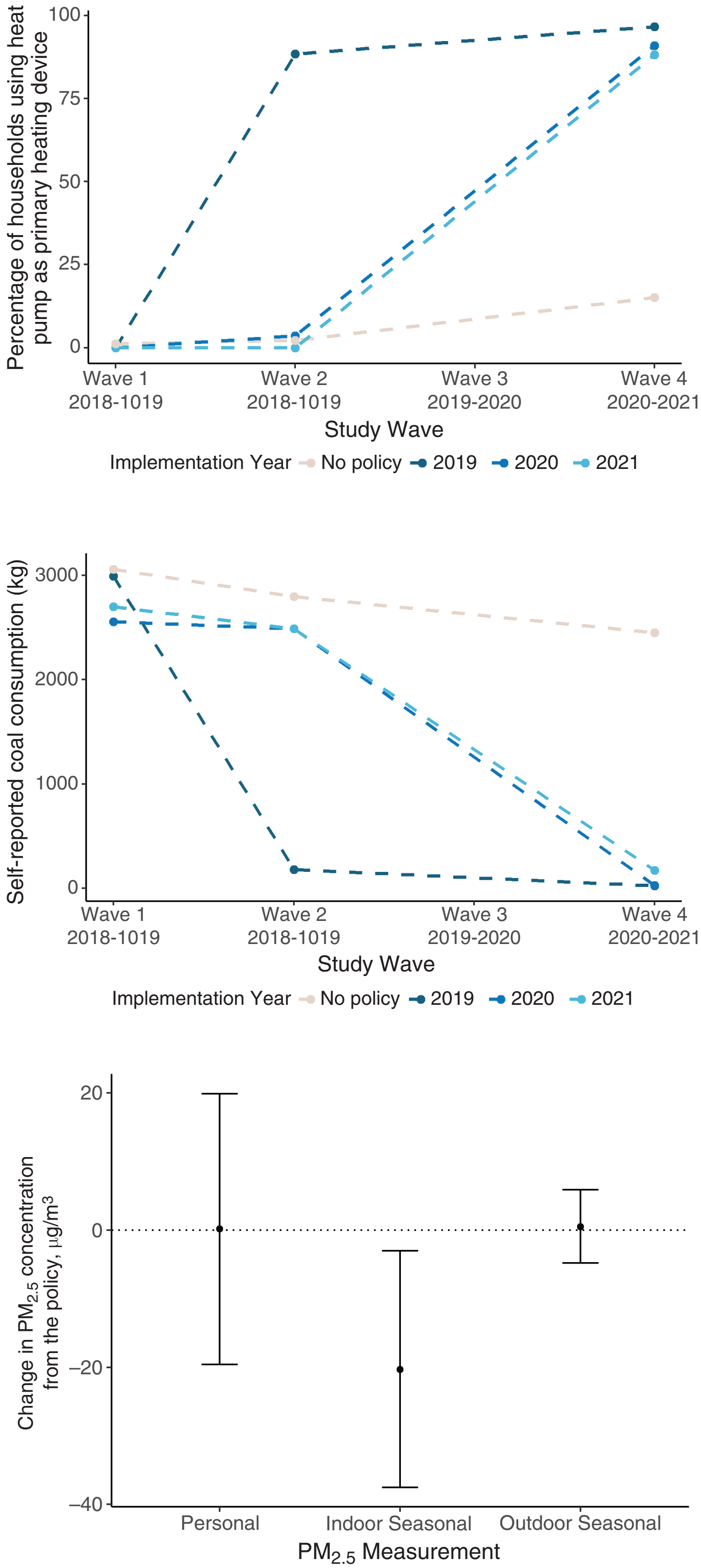
Among villages where the policy was not implemented by the end of the study, the proportion of households that reported using electric heat pumps increased from 3% at the start of the study to 16% at the end of the study, and most of those households that used heat pumps reported using them exclusively.

Implementation of the policy corresponded with steep declines in the use of self-reported coal use (**Commentary Figure 3, center**). Self-reported electricity expenditures increased over time, regardless of whether the villages had implemented the policy. At the same time, self-reported use of biomass — which was not targeted by the policy — remained mostly stable.

EFFECTS OF THE POLICY ON AIR QUALITY AND TEMPERATURE

Measured personal concentrations of $PM_{2.5}$ and black carbon were higher than measured indoor concentrations, which were higher than measured outdoor concentrations. This was the case both for villages where the clean heating policy was implemented and where it was not. Generally, personal, indoor, and outdoor concentrations decreased over the four winters of the study period, regardless of policy implementation. For example, seasonal average $PM_{2.5}$ concentrations over all villages decreased from 38 µg/m³ in Wave 1 to 26 µg/m³ in Wave 4. Baseline personal exposures to $PM_{2.5}$ and black carbon were 20% and 2% lower, respectively, in villages where the policy was eventually implemented than in other villages, which could mask any effects of the policy. However, the investigators were able to account for this difference at baseline using the difference-in-differences approach.

Implementation of the policy reduced indoor $PM_{2.5}$ concentrations by about 20 µg/m³ but had no effect on outdoor and personal $PM_{2.5}$ concentrations (**Commentary Figure 3, bottom**). The effect on indoor $PM_{2.5}$ concentrations was roughly one quarter of the baseline Wave 1 concentrations in 2018–2019 and about the same as the average change in indoor $PM_{2.5}$ concentrations across all residences between Wave 2 (2019–2020) and Wave 4 (2021–2022).



Commentary Figure 3. Effects of the policy on heat pump availability, coal consumption, and air quality. Top: Percentage of households using heat pumps as primary heating device. Center: Self-reported household coal consumption in kg. Darker colors represent villages where the clean heating policy was implemented earlier. Points indicate when the questionnaires were completed; effects on villages where the policy was implemented in 2019 or 2020 were first observed in the 2021 data collection because no questionnaires were collected in Wave 3. Bottom: Effect of the clean heating policy on personal, indoor, and outdoor $PM_{2.5}$ concentrations with 95% confidence intervals, adjusted for household size, tobacco smoking category, outdoor temperature, and outdoor dew point (negative values mean that the policy reduced $PM_{2.5}$).

From their source apportionment analyses, the investigators identified a factor that they called “mixed combustion” because it included chemical species typical of direct emissions from coal combustion, biomass burning, and possibly tobacco smoking. Over the study period, “mixed combustion” accounted for about 5–10 μg/m^3 of the outdoor $PM_{2.5}$ and about 20–30 μg/m^3 of the personal $PM_{2.5}$. Implementation of the policy had statistically insignificant and imprecise effects on “mixed combustion” exposures.

Households in villages that had implemented the policy had about 2°C higher mean and 4°C higher minimum winter indoor temperatures than households in villages without the policy, even though both groups of households had similar winter indoor temperatures at baseline. The investigators did not report any systematic heterogeneity by the year of policy implementation in the effect of the policy on $PM_{2.5}$ concentrations or temperature.

TOTAL EFFECTS OF THE POLICY ON HEALTH OUTCOMES

Individuals who lived in households in villages where the policy had been implemented had improved blood pressure and self-reported chronic respiratory symptoms. In adjusted models, both systolic and diastolic blood pressure decreased by about 1.5 mm Hg after villages implemented the policy, although there was no meaningful change in pulse pressure or blood pressure amplification. For comparison, the literature suggests that each 5 mm Hg reduction in systolic blood pressure reduces the risk of developing cardiovascular events by 10% (Canoy et al. 2022).

Implementation of the policy was associated with a 7.5% reduction in self-reported respiratory symptoms. This finding was largely driven by reductions in reports of coughing, having chest trouble, or difficulty breathing on several or most days of the week. The investigators did not find evidence that exposure to the policy affected phlegm, wheezing symptoms, levels of biomarkers in blood, or levels of FeNO in exhaled air.

Effects of the policy on blood pressure and phlegm were strongest in the villages where the policy was implemented earliest and were weaker in later years. The investigators did not report systematic patterns related to when the policy was implemented for other respiratory symptoms.

DIRECT AND INDIRECT EFFECTS OF THE POLICY ON HEALTH OUTCOMES

Because the overall models provided evidence that the policy had affected blood pressure and self-reported respiratory symptoms, the investigators conducted mediation analyses on whether changes to these health outcomes could be attributed to changes in $PM_{2.5}$ and temperature (**Commentary Table 2**). In the mediation analyses, the estimates of direct effects of the policy on all four measures of blood pressure decreased, suggesting that the policy had its effects on health mediated through changes in air quality and temperature. When the mediation of the effect of the policy on blood pressure by $PM_{2.5}$ and indoor temperature together was tested, the direct effect of the policy almost completely disappeared, suggesting that the effect of the policy on blood pressure could be explained almost completely by the policy-related reductions in indoor $PM_{2.5}$ concentrations and indoor temperature.

The direct effects for the self-reported respiratory outcomes were statistically indistinguishable from the total effects, so the investigators concluded that changes in $PM_{2.5}$ and temperature were not contributing factors to the policy-related changes in respiratory outcomes.

HEI REVIEW COMMITTEE EVALUATION

Baumgartner, Harper, and colleagues evaluated the air quality and health effects of China’s policy to replace coal heaters with heat pumps in 50 villages that did not have the clean heating policy at the start of the study. By the end of the study, the policy had been implemented in 20 of the villages, and there was nearly complete compliance with the policy. Implementation of the policy reduced indoor $PM_{2.5}$ by about 20 μg/m^3, with smaller or negligible changes in outdoor $PM_{2.5}$, personal $PM_{2.5}$, and personal black carbon. At the same time, indoor temperatures during winter increased after implementation of the policy. The policy also slightly improved blood pressure and respiratory symptoms. Almost all the effects of the policy on blood pressure could be explained by the imputed improvements in indoor or personal $PM_{2.5}$ concentrations and indoor temperature, and the improvements in respiratory symptoms were most likely because of other factors.

In its independent evaluation of the study, the Review Committee thought that Baumgartner, Harper, and colleagues had completed an important study to evaluate the benefits of a clean energy policy on air quality and health. The study design and the causal framework in which they applied DAGs, staggered difference-in-differences analyses, and mediation analyses were strong. The Committee was impressed with the efforts to adapt the study design to the least extent necessary during the COVID-19 pandemic. Other strengths of the study included the use of sensors alongside filter samplers to measure $PM_{2.5}$ and long-term sampling in rural areas. Additionally, the statistical analysis methods were well described, and the extended two-way fixed effects model was well justified with details on the model and the DAG. The Committee agreed with the overall conclusion that the clean heating policy achieved its intended goals by dramatically reducing residential coal burning and improving indoor environmental quality, which provided some benefits to blood pressure and respiratory symptoms.

DIFFERENCE-IN-DIFFERENCES MODELING

The Review Committee highlighted that the staggered difference-in-differences modeling was generally a strong approach because it accounted for several key sources of potential confounding. First, this design removed all temporally fixed differences in the study populations in villages where the policy was implemented and where it was not. It

Commentary Table 2. Summary of Mediation Analysis Results[a]

	Was this factor found to substantially mediate the effect of the policy? (Percentage of the total effect that could be explained by mediation, ranges for different health outcomes in this group)			
Health Outcome or Effect	**Indoor or Personal $PM_{2.5}$[b]**	**Indoor Temperature[c]**	**Indoor or Personal $PM_{2.5}$ and Indoor Temperature[b,c]**	**"Mixed Combustion"**
Measured brachial blood pressure (systolic or diastolic)	YES (31% to 43%)	YES (31% to 79%)	YES (63% to 121%)	Not tested
Measured central blood pressure (systolic or diastolic)	YES (31% to 43%)	YES (25% to 71%)	YES (56% to 114%)	NO (–44% to 6%)
Self-reported respiratory outcomes (any symptom, coughing, phlegm, wheezing attacks, trouble breathing, or chest trouble)	NO (–13% to 33%)	NO (–143% to 8%)	NO (–40% to 15%)	Not tested

[a] Only those outcomes where mediation analyses were conducted are included in the table. No mediation analyses were conducted for pulse pressure, blood pressure amplification, FeNO, or inflammatory blood markers because there was no overall effect of the policy on these outcomes.

[b] Indoor $PM_{2.5}$ concentrations from about one-third of the participants in Waves 2–4 were used for blood pressure, and personal $PM_{2.5}$ concentrations from half of the participants in Waves 1, 2, and 4 were used for respiratory effects.

[c] Seasonal average temperatures were used for blood pressure, and 5-minute average temperatures were used for respiratory outcomes.

also accounted for background trends in air pollution and health during the study that might have occurred across the area regardless of whether the policy had been implemented. Nonetheless, some limitations in the difference-in-differences models and how they were used by the investigators bear discussion.

The key assumption for the difference-in-differences approach was that the time trends in the villages without the policy are an accurate reflection of the time trends that would have occurred in the villages with the policy had they, counterfactually, not implemented the policy. Because this policy was implemented over a period spanning years before, during, and after the COVID-19 pandemic, one must consider if the study conduct, the intervention implementation, or the outcomes could have been affected differently in the villages with and without the policy. There are several assurances that this is not the case. First, the investigators held to the original study plan wherever possible while also adjusting as needed. For example, they continued village-level tracking of the policies even when they were unable to recruit new participants or collect individual-level data. Second, the investigators also confirmed, to the extent possible, that all villages experienced similar COVID-19 burdens. In fact, according to the best available data, there were no COVID-19 cases in any of the suburban and rural villages of the investigation during the study period. Additionally, the investigators report that the villages all had similar requirements to prevent the spread of COVID-19. Collectively, these factors provided confidence to the investigators and the Committee that there should be similar trends in the health outcomes and exposures across villages with and without interventions during the pandemic period.

The investigators also explored the possibility of other unexplained differences in the trends of air pollution exposure and health outcomes by comparing pre-policy trends in the villages that implemented the policy at different times during the study period versus those that did not implement the policy. Largely, they observed no differences in trends between the villages before the policy implementation. The only exposure with different trends between villages with and without the policy was personal $PM_{2.5}$, which was found to have higher baseline concentrations and to be initially declining more steeply in villages where the policy was never implemented, as compared to those where the policy was eventually implemented. Similarly, there was a greater development of the probability of cough and chest complaints in the villages that received the intervention in 2021 as compared to other groups. For all situations, however, these trends suggest that the effect of the policy, if anything, would be underestimated.

Overall, the Committee thought that the investigators had carefully applied difference-in-differences modeling, that they addressed the limitations of the approach in a real-world accountability study, and that the results were robust.

CHALLENGES FOR MEDIATION ANALYSIS USING IMPUTED DATA

The use of causal mediation analysis was also felt to be a constructive addition to the study. At the same time, a major challenge for the mediation analysis was that data were needed on parameters that had little overlap in availability, thus leading the investigators to impute data for parameters that were not measured. In particular, it should be noted that imputation was used to account for missing indoor $PM_{2.5}$ measurements for the majority of participants, which adds uncertainty to analyses that involved indoor $PM_{2.5}$. The use of imputation might have made it difficult to detect true mediation or introduced measurement error or noise to the analyses. As a result, the Committee thought that the results of analyses using imputed data, especially the mediation analyses for blood pressure, should be treated as exploratory and interpreted with caution.

OPPORTUNITIES FOR FUTURE ANALYSES

Some of the surprising results in this study suggest the opportunity for future analyses that were out of the scope of the current study but might provide additional insight into the effects of the clean heating policy and the mediating factors through which it works.

Surprisingly, the investigators did not find that the clean heating policy had significant effects on personal or outdoor seasonal $PM_{2.5}$. It is possible that some of this difference had to do with averaging across an entire season when the intensity of heating (and thus fuel combustion) might have varied throughout the season. Therefore, it is possible that the policy did affect personal and outdoor $PM_{2.5}$ even though no effect was found over the entire winter season. Future analyses making use of heating-degree-days might be able to test for potential effects of the policy on $PM_{2.5}$ on the days where the policy would be expected to affect emissions the most.

Because less than one-third of the total effect of the policy on respiratory symptoms was mediated by indoor $PM_{2.5}$, a more complete understanding of the effect of the policy on respiratory health might be obtained if different potential mediators were tested, such as other pollutants, time spent cooking, heating days, and thermal comfort.

PUTTING THE AIR POLLUTANT EXPOSURES INTO GLOBAL CONTEXT

To put the air pollution exposures measured in the study into context, the investigators compared the measured baseline personal $PM_{2.5}$ concentrations to WHO air quality guidelines, which are recommendations for health-protective concentrations of ambient air pollution (WHO 2021). Assuming that the measured wintertime personal concentrations were similar to annual average ambient concentrations, which is a large assumption, the investigators stated that the observed personal $PM_{2.5}$ concentrations were aligned with WHO's Interim Target 1. The Committee observed that the seasonal average outdoor $PM_{2.5}$ concentrations were also of similar magnitude. They agreed with the investigators that $PM_{2.5}$ concentrations — whether personal, indoor, or outdoor — in this region were high. It was therefore valuable to evaluate the effect of China's clean heating policy on air pollution and health, and reassuring to see that the policy had some effects.

SUMMARY AND CONCLUSIONS

Baumgartner, Harper, and colleagues conducted a thorough accountability study of the air quality and health effects of China's clean heating policy. They evaluated more than 1,400 participants from more than 1,200 households in 50 suburban and rural villages of Beijing over four consecutive winters between 2018 and 2022. By the end of the study period, 20 villages had implemented the policy, which led to improved air quality and some health improvements in individuals living in those villages. The improvements in blood pressure were linked to improvements in air quality and indoor temperature.

The HEI Review Committee members were impressed with the strong study design and causal framework, showing that the policy was linked to some improvements in air quality that were in turn linked to some health improvements. They appreciated the difference-in-differences approach, which allowed the study team to account for many factors other than the clean heating policy that could have resulted in differences between the villages or changes in air pollution and health over time. They also acknowledged the effort involved for the investigators to follow the original study plan wherever possible, including through the COVID-19 pandemic.

Overall, the study demonstrated that the clean heating policy achieved its intended goals to electrify household heating in the villages where it was implemented and that it dramatically reduced residential coal burning and improved indoor environmental quality in the first years after implementation. The policy provided some benefits to heart and lung health, some of which (systolic and diastolic blood pressure) were related to decreases in air pollution exposure.

Although there is an abundance of evidence on associations of negative health outcomes with exposure to $PM_{2.5}$, it is important to quantify (rather than assume) the effects of specific actions on air quality and health. This study provides evidence that replacing coal-fueled heaters with heat pumps reduces indoor air pollution and improves health. These results are encouraging for other countries seeking to implement policies to replace highly polluting residential heating sources.

ACKNOWLEDGMENTS

The HEI Review Committee thanks the ad hoc reviewers for their help in evaluating the scientific merit of the Investigators' Report. The Committee is also grateful to Hanna Boogaard and Eleane van Vliet for oversight of the study, to Allison Patton for assistance with review of the Investigators' Report and preparation of its Commentary, to Progressive Publishing Services and Tom Zaczkiewicz for editing the Investigators' Report and its Commentary, and to Kristin Eckles for her role in preparing this Research Report for publication.

REFERENCES

Barrington-Leigh C, Baumgartner J, Carter E, Robinson BE, Tao S, Zhang Y. 2019. An evaluation of air quality, home heating, and well-being under Beijing's programme to eliminate household coal use. Nat Energy 4:416–423; *https://doi.org/10.1097/01.EE9.0000606212.67011.88*.

Canoy D, Nazarzadeh M, Copland E, Bidel Z, Rao S, Li Y, et al. 2022. How much lowering of blood pressure is required to prevent cardiovascular disease in patients with and without previous cardiovascular disease? Curr Cardiol Rep 24:851–860; *https://doi.org/10.1007/s11886-022-01706-4*.

China State Council. 2013. Air Pollution Prevention and Control Action Plan: China Clean Air Updates (English Translation). Clean Air Alliance of China. Available: *https://policy.asiapacificenergy.org/sites/default/files/Air%20Pollution%20Prevention%20and%20Control%20Action%20Plan%20%28EN%29.pdf* [accessed 8 August 2025].

Chowdhury S, Pillarisetti A, Oberholzer A, Jetter J, Mitchell J, Cappuccilli E, et al. 2023. A global review of the state of the evidence of household air pollution's contribution to ambient fine particulate matter and their related health impacts. Environment International 173:107835; *https://doi.org/10.1016/j.envint.2023.107835*.

Ding D, Xing J, Wang S, Liu K, Hao J. 2019. Estimated contributions of emissions controls, meteorological factors, population growth, and changes in baseline mortality to reductions in ambient $PM_{2.5}$ and $PM_{2.5}$-related mortality in China, 2013–2017. Environ Health Perspect 127:067009; *https://doi.org/10.1289/ehp4157*.

Dispersed Coal Management Research Group. 2023. 中国散煤综合治理研究报告 (China Dispersed Coal Governance Report). 北京大学能源研究院气候变化与能源转型项目 (Climate Change and Energy Transition Project, Institute of Energy, Peking University). Available: *https://www.ditan.com/static/upload/file/20231106/1699241777630902.pdf* [accessed 23 July 2025].

GBD MAPS Working Group. 2016. Burden of Disease Attributable to Coal-Burning and Other Air Pollution Sources in China. Health Effects Institute. Available: *https://www.healtheffects.org/publication/burden-disease-attributable-coal-burning-and-other-air-pollution-sources-china* [accessed 22 July 2025].

Health Effects Institute (HEI). 2025. State of Global Air 2025: A Report on Air Pollution and Its Role in the World's Leading Causes of Death. Boston, MA: Health Effects Institute.

HEI Household Air Pollution Working Group. 2018. Household Air Pollution and Noncommunicable Disease. Communication 18. Boston, MA: Health Effects Institute.

Hernán M, Robins J. 2020. Causal Inference: What If. Boca Raton, FL: Chapman & Hall/CRC.

IEA, IRENA, UNSD, World Bank, WHO. 2025. Tracking SDG 7: The Energy Progress Report. Washington, DC: World Bank. Available: *https://www.irena.org/Publications/2024/Jun/Tracking-SDG-7-The-Energy-Progress-Report-2024*.

Keele L, Tingley D, Yamamoto T. 2015. Identifying mechanisms behind policy interventions via causal mediation analysis: Methods for policy analysis. J Pol Anal Manage 34:937–963; *https://doi.org/10.1002/pam.21853?urlappend=%3Futm_source%3Dresearchgate.net%26medium%3Darticle*.

Kinney P, Beizhan Yan, Xiaoming Shi, Maigeng Zhou, Shuxiao Wang, Haotian Zheng, et al. In press. Accounting for the Health Benefits of Air Pollution Regulations in China, 2008–2019. Research Report. Health Effects Institute: Boston, MA.

Lai PS, Lam NL, Gallery B, Lee AG, Adair-Rohani H, Alexander D, et al. 2024. 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 209:909–927; *https://doi.org/10.1164/rccm.202402-0398st*.

Lee KK, Bing R, Kiang J, Bashir S, Spath N, Stelzle D, et al. 2020. Adverse health effects associated with household air pollution: A systematic review, meta-analysis, and burden estimation study. Lancet Glob Health 8:e1427–e1434; *https://doi.org/10.1016/s2214-109x(20)30343-0*.

Lee M, Chang J, Deng Q, Hu P, Bixby H, Harper S, et al. 2024. Effects of a coal-to-clean heating policy on acute myocardial infarction in Beijing: A difference-in-differences analysis. Lancet Planet Health 8:e924–e932; *https://doi.org/10.1016/s2542-5196(24)00243-2*.

Li T, Chen C, Zhang M, Zhao L, Liu Y, Guo Y, et al. 2024. Accountability analysis of health benefits related to National Action Plan on Air Pollution Prevention and Control in China. PNAS Nexus 3:pgae142; *https://doi.org/10.1093/pnasnexus/pgae142*.

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

Wooldridge JM. 2021. Two-way fixed effects, the two-way Mundlak regression, and difference-in-differences estimators. SSRN (August 17, 2021); *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3906345*.

World Health Organization (WHO). 2014. WHO Guidelines for Indoor Air Quality: Household Fuel Combustion. Geneva, Switzerland: WHO. Available: *https://www.who.int/publications/i/item/9789241548885*.

World Health Organization (WHO). 2021. WHO Global Air Quality Guidelines: Particulate Matter ($PM_{2.5}$ and PM_{10}), Ozone, Nitrogen Dioxide, Sulfur Dioxide, and Carbon Monoxide. Available: *https://www.who.int/publications/i/item/9789240034228* [accessed 21 July 2025].

Zhang R, Jing J, Tao J, Hsu S-C, Wang G, Cao J, et al. 2013. Chemical characterization and source apportionment of $PM_{2.5}$ in Beijing: Seasonal perspective. Atmos Chem Phys 13:7053–7074; *https://doi.org/10.5194/acpd-13-9953-2013*.

Zhang Z, Wang W, Cheng M, Liu S, Xu J, He Y, et al. 2017. The contribution of residential coal combustion to $PM_{2.5}$ pollution over China's Beijing-Tianjin-Hebei region in winter. Atmos Environ 159:147–161; *https://doi.org/10.1016/j.atmosenv.2017.03.054*.

Zhang Q, Zheng Y, Tong D, Shao M, Wang S, Zhang Y, et al. 2019. Drivers of improved $PM_{2.5}$ air quality in China from 2013 to 2017. Proc Natl Acad Sci USA 116:24463–24469; *https://doi.org/10.1073/pnas.1907956116*.

ABBREVIATIONS AND OTHER TERMS

ATT	average treatment effect on the treated
BC	black carbon
BP	blood pressure
BS_mapping	bootstrap mapping
CI	confidence interval
CIE	International Commission on Illumination
CHP	Clean Heating Policy
cDBP	central diastolic blood pressure
CRP	C-reactive protein
cSBP	central systolic blood pressure
DAG	directed acyclic graph
DiD	difference-in-differences
DISP	displacement of factor elements
DISP%	displacement percentage
EC	elemental carbon
EDXRF	evo energy-dispersive x-ray fluorescence
ETWFE	extended two-way fixed effects
FEM	federal equivalent method
FID	flame ionization detector
FeNO	fractional exhaled nitric oxide
GBD MAPS	The Global Burden of Disease — Major Air Pollution Sources
HAPIN	Household Air Pollution Intervention Network
HPLC	high-performance liquid chromatography
IL-6	interleukin-6
MDA	malondialdehyde
NIOSH	National Institute for Occupational Safety and Health
NISP	National Improved Stove Program
NIST	National Institute of Standards and Technology
ns-S	nonsulfate sulfur
OC	organic carbon
OD	optic densities
$PM_{2.5}$	particulate matter ≤2.5 µm in aerodynamic diameter
Qexp	expected fit
Qtrue	raw goodness-of-fit statistic (Σ scaled residual^2) calculated with all observations included
Qrobust	goodness-of-fit statistic recalculated after the algorithm automatically down-weights any data points whose uncertainty-scaled residual $\lvert e/\sigma \rvert > 4$
Qr/Qexp	ratio of the robust fit to the statistically expected Q (≈ number of valid observations – number of free parameters)
Q/Qexp	model fit divided by expected fit
RMSE	root mean square error
SRM	standard reference material
TNF-α	tumor necrosis factor-alpha
UPAS	ultrasonic personal aerosol samplers
wi	water-insoluble species
ws	water-soluble species

RELATED HEI PUBLICATIONS

Number	Title	Principal Investigator	Date
Research Reports			
224	Effect of Air Pollution Reductions on Mortality During the COVID-19 Lockdowns in Early 2020	K. Chen	2025
223	Impacts of Vehicle Emission Regulations and Local Congestion Policies on Birth Outcomes Associated with Traffic Air Pollution	P. Hystad	2025
221	Assessing the National Health, Education, and Air Quality Benefits of the United States Environmental Protection Agency's School Bus Rebate Program: A Randomized Controlled Trial Design	S.D. Adar	2024
218	Estimating Model-Based Marginal Societal Health Benefits of Air Pollution Emission Reductions in the United States and Canada	A. Hakami	2024
205	Improvements in Air Quality and Health Outcomes Among California Medicaid Enrollees Due to Goods Movement Actions	Y-Y. Meng	2021
199	Real-World Vehicle Emissions Characterization for the Shing Mun Tunnel in Hong Kong and the Fort McHenry Tunnel in the United States	X. Wang	2019
195	Impacts of Regulations on Air Quality and Emergency Department Visits in the Atlanta Metropolitan Area, 1999–2013	A. Russell	2018
190	The Effects of Policy-Driven Air Quality Improvements on Children's Respiratory Health	F. Gilliland	2017
189	Ambient Air Pollution and Adverse Pregnancy Outcomes in Wuhan, China	Z. Qian	2016
188	Adverse Reproductive Health Outcomes and Exposure to Gaseous and Particulate-Matter Air Pollution in Pregnant Women	J. Wu	2016
187	Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations	C.M. Zigler	2016
183	Development of Statistical Methods for Multipollutant Research		
	Part 1. Statistical Learning Methods for the Effects of Multiple Air Pollution Constituents	B.A. Coull	2015
	Part 2. Development of Enhanced Statistical Methods for Assessing Health Effects Associated with an Unknown Number of Major Sources of Multiple Air Pollutants	E.S. Park	2015
	Part 3. Modeling of Multipollutant Profiles and Spatially Varying Health Effects with Applications to Indicators of Adverse Birth Outcomes	J. Molitor	2016
Special Reports			
20	Burden of Disease Attributable to Coal-Burning and Other Air Pollution Sources in China. Health Effects Institute	GBD MAPS Working Group	2016
	State of Global Air 2025: A Report on Air Pollution and Its Role in the World's Leading Causes of Death	HEI	2025
HEI Communications			
18	Household Air Pollution and Noncommunicable Disease	HEI Household Air Pollution Working Group	2018

PDFs are available for free downloading at www.healtheffects.org/publications.

HEI BOARD, COMMITTEES, and STAFF

BOARD OF DIRECTORS

RESEARCH COMMITTEE

continued on next page

HEI BOARD, COMMITTEES, and STAFF

REVIEW COMMITTEE

Sara D. Adar *Professor of Epidemiology and Global Public Health, Department of Epidemiology, University of Michigan School of Public Health, USA*

Katherine B. Ensor *Noah G. Harding Professor of Statistics, Rice University, USA*

Ulrike Gehring *Associate Professor, Institute for Risk Assessment Sciences, Utrecht University, Netherlands*

Michael Jerrett *Professor, Department of Environmental Health Sciences, Fielding School of Public Health, University of California Los Angeles, USA*

John Volckens *Professor, Department of Mechanical Engineering, Walter Scott Jr. College of Engineering, Colorado State University, USA*

Scott Weichenthal *Professor, Department of Epidemiology, Biostatistics, and Occupational Health, School of Population and Global Health, McGill University, Canada*

STAFF AND CONSULTING SCIENTISTS

Elena Craft *President and CEO*

Ellen K. Mantus *Director of Science*

Donna J. Vorhees *Director of HEI Energy*

Thomas J. Champoux *Director of Science Communications*

Jacqueline C. Rutledge *Director of Finance and Administration*

Emily Alden *Corporate Secretary*

Daniel S. Greenbaum *President Emeritus, In Memoriam 1952–2024*

Robert M. O'Keefe *Vice President Emeritus*

Annemoon M. van Erp *Deputy Director of Science Emerita*

Amy Andreini *Science Communications Specialist*

Hanna Boogaard *Consulting Principal Scientist*

Jacki Collins *Senior Staff Accountant*

Dan Crouse *Senior Scientist*

Cloelle Danforth *Senior Scientist*

Philip J. DeMarco *Compliance Manager*

Kristin C. Eckles *Senior Editorial Manager*

Hlina Kiros *Research Assistant*

Lissa McBurney *Senior Science Administrator*

Samantha Miller *Research Assistant*

Victor Nthusi *Consulting Research Fellow*

Pallavi Pant *Head of Global Initiatives*

Allison P. Patton *Senior Scientist*

Yasmin Romitti *Staff Scientist*

Anna Rosofsky *Senior Scientist and Community Health and Environmental Research Initiatives Lead*

Continues next page

RESEARCH REPORT

NUMBER 235

DECEMBER 2025

Health Effects Institute

75 Federal Street
Suite 1400
Boston, Massachusetts 02110, USA
+1-617-488-2300

www.healtheffects.org