ARTICLE OPEN

Check for updates

Source apportionment of fine particulate matter ($PM_{2.5}$) personal exposures: findings from the Household Air Pollution Intervention Network (HAPIN) study in rural Guatemala

Erick Mollinedo[1], Katherine A. Kearns[1,2], Armistead G. Russell[3], Michael Johnson[2], Christian L'Orange[4], Ricardo Piedrahita[2], Ajay Pillarisetti[5], Jeremy A. Sarnat[6], John P. McCracken[7,8], Lisa M. Thompson[9], Anaité Diaz-Artiga[8], Maggie L. Clark[10], Kyle Steenland[6], Lance A. Waller[11], Thomas F. Clasen[6], Jennifer Peel[10], William Checkley[12], Luke P. Naeher[1]✉ and HAPIN Investigators*

© The Author(s) 2026

BACKGROUND: Source apportionment is a method that reconstructs sources of pollution from monitored values. The identification of these sources is important to develop interventions and other strategies to improve environmental quality.
OBJECTIVE: This study aimed to identify and quantify potential sources that contribute to fine particulate matter ($PM_{2.5}$) personal exposures in a cohort of pregnant women participating in a fuel-cooking intervention trial in Guatemala.
METHODS: We estimated the $PM_{2.5}$ and black carbon (BC) concentrations from 629 polytetrafluoroethylene (PTFE) filter samples using gravimetric and transmittance analyses, and we analyzed inorganic elements using energy dispersive X-ray fluorescence (ED-XRF). The filters correspond to personal exposure samples collected with the Enhanced Children's MicroPEM™ (ECM) from pregnant women who cook using biomass or liquefied petroleum gas (LPG) fuels within the Household Air Pollution Intervention Network (HAPIN) randomized-controlled trial in rural Jalapa, Guatemala. We used the U.S. Environmental Protection Agency's (EPA) Positive Matrix Factorization (PMF) model to identify the potential sources.
RESULTS: A four-factor source apportionment model was derived from PMF. The potential sources and their relative contributions to $PM_{2.5}$ mass were identified as: biomass burning (~69.6%), fossil fuels (22%), crustal, and other soil (~4% each). When categorizing our samples by study group, baseline (155 μg/m^3; 95% CI: 129.7, 180.2) and post-intervention control (127.2 μg/m^3; 95% CI: 114.7, 139.7) samples had the highest amount of $PM_{2.5}$ mass (49.7% and 40.7%, respectively) compared to 9.6% (29.9 μg/m^3; 95% CI: 27.2, 32.7) from intervention samples.
SIGNIFICANCE: We identified biomass burning, fossil fuel burning, crustal and other soil as sources of $PM_{2.5}$ pollution in rural Jalapa, Guatemala. These findings pave the road for future source apportionment studies in this region and highlight the importance of source characterization to implement interventions to reduce $PM_{2.5}$ emissions.
IMPACT: We identified four potential sources of $PM_{2.5}$ in rural Jalapa, Guatemala, based on personal exposure samples from pregnant women participating in the HAPIN trial. Biomass and fossil fuel burning are the sources with the highest contributions, followed by crustal and other soil. The findings of this study highlight the potential to prioritize emissions reduction efforts in rural Guatemala through the implementation of specific interventions based on the derived sources of pollution.

Keywords: Household air pollution; Exposure assessment; Fine particulate matter; Source apportionment; Positive matrix factorization; Guatemala

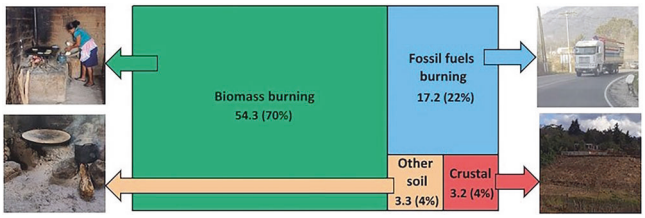
Journal of Exposure Science & Environmental Epidemiology; https://doi.org/10.1038/s41370-026-00910-6

[1]Department of Environmental Health Sciences, College of Public Health, University of Georgia, Athens, GA, USA. [2]Berkeley Air Monitoring Group, Berkeley, CA, USA. [3]School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, USA. [4]Department of Mechanical Engineering, Colorado State University, Fort Collins, CO, USA. [5]Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA, USA. [6]Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA. [7]Global Health Institute, College of Public Health, University of Georgia, Athens, GA, USA. [8]Center for Health Studies, Universidad del Valle de Guatemala, Guatemala City, Guatemala. [9]Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, GA, USA. [10]Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, CO, USA. [11]Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, USA. [12]Division of Pulmonary and Critical Care, School of Medicine, Johns Hopkins University, Baltimore, MD, USA. *A list of authors and their affiliations appears at the end of the paper. ✉email: lnaeher@uga.edu

Received: 4 June 2025 Revised: 12 April 2026 Accepted: 16 April 2026
Published online: 26 June 2026

Graphical Abstract

Sources and average $PM_{2.5}$ mass contribution in $\mu g/m^3$ (%) from rural Jalapa, Guatemala.



INTRODUCTION

Household air pollution (HAP) from the combustion of biomass for cooking is among the most prominent environmental health issues in low- and middle-income countries (LMICs) [1]. In 2022, it was estimated that around 2.1 billion people relied on biomass for cooking [2]. HAP, which includes ambient and indoor particulate and gaseous air pollutant components, is the leading environmental risk factor contributing to the global burden of disease [3]. The resulting exposure to particulate household air pollution was linked to more than 3.1 million deaths in 2021 [3, 4]. HAP is associated with adverse cardiovascular and respiratory diseases (e.g., chronic obstructive pulmonary disease, ischemic heart disease, lower respiratory infections, and lung cancer), low birthweight, and preterm birth [1, 5–7]. Fine particulate matter with an aerodynamic diameter less than or equal to 2.5 micrometers ($PM_{2.5}$) is one of the major constituents of HAP. $PM_{2.5}$ is chemically and physically heterogeneous and composed largely of black carbon (BC), organic carbon (OC), ions, and organic and inorganic elements and compounds, and can originate from multiple natural and anthropogenic sources [8, 9].

The identification of sources of pollution is important for the development of control strategies and policies aiming to improve air quality [10, 11]. Source apportionment (SA) reconstructs potential emissions' sources from mixtures of pollutants or individual pollutants [12]. The top-down receptor-based source apportionment method involves conducting ambient sampling, performing chemical analysis on the samples, and then using receptor models to derive potential sources [13, 14]. Receptor models are mathematical packages that quantify the contribution of sources to samples based on the profiles of sources of either known or unknown identities [15]. Positive matrix factorization (PMF), a widely used SA top-down receptor model, decomposes the environmental samples into a matrix of factor contributions and a matrix of factor profiles. It works by grouping the components from this decomposition as clusters, and the correlated clusters are categorized as profiles. Then, these profiles are identified as potential sources based on pre-existing information from the literature. PMF forces the contributions to be non-negative, and incorporates the instrument's or user-provided uncertainties to account for the measurement's confidence, and weighs in the influence of the individual measurements on the solution [15–17]. One advantage of PMF over other receptor models is that it does not require prior knowledge of specific information in a specific location about source emissions and their profiles [15, 18].

PMF was first used for source apportionment studies in the United States [19–24], but more recently it has expanded globally, including Mexico [25–27] and countries in Asia [28–32]. In South America, a few studies have been conducted in Colombia [33], Brazil [34, 35] and Chile [36, 37]. However, these studies evaluate the sources of ambient air pollution, instead of from personal exposure samples, and most of them are conducted in heavily populated or urban areas. One study in Ghana [38] and another in India [39] used PMF to highlight the contribution of biomass burning as a source of ambient $PM_{2.5}$. In terms of HAP in LMICs, few studies have evaluated the composition of $PM_{2.5}$ from household air pollution samples, one in Ghana and Gambia [40] and another in China [41]. Both studies determined that biomass burning was one of the highest source contributors to total $PM_{2.5}$ mass.

Given that, to the authors' knowledge, there are no source apportionment studies in Central America [11], and there are a limited number of studies in rural areas where HAP is pervasive, we aimed to characterize the sources of $PM_{2.5}$ in rural Guatemala and to quantify the contributions of $PM_{2.5}$ by source. This research is anchored to the Household Air Pollution Intervention Network (HAPIN) trial [42–44], a randomized controlled trial in four low- and middle-income countries that aimed to evaluate the effect of a liquefied petroleum gas (LPG) intervention on exposure to HAP and health.

METHODS

Study setting

Particulate matter sampling was conducted in rural communities of Santa Maria Xalapam from the Jalapa municipality in the Jalapa department of Guatemala. There are around 159,840 inhabitants in the municipality of Jalapa, and around 66% of the population relies on biomass fuel (mainly wood) for cooking [45]. This site is located between 1500 and 2000 meters above sea level [46]. The climate is mild and temperate, characterized by a rainy season and a dry season [42].

Study design

Personal $PM_{2.5}$ exposure samples were collected from pregnant women participating in the HAPIN randomized controlled trial at the Guatemala site. In summary, the intervention consisted of a free LPG cookstove, a continuous supply of free LPG fuel to the participant's homes until the infant was 1 year of age, and educational and behavioral messaging for the safe and exclusive use of LPG. In Guatemala, the LPG stove consisted of a small burner, a medium burner, and a large burner with a removable metal plate (comal). $PM_{2.5}$ was collected on polytetrafluoroethylene (PTFE) filters with a diameter of 15 mm and a pore size of 2 µm (Measurement Technology Laboratories, Minnesota, USA). Filters were used in the Enhanced Children's MicroPEM™ (ECM, RTI International, Research Triangle Park, NC, USA). Broadly, participants were fitted with an apron with a pocket designed to carry an ECM, which was programmed to sample for 24 hours with a flow rate of 0.3 liters per minute. In total, 630 samples and 18 blank filters were randomly selected to be representative of all HAPIN pregnancy visits. Blank filter samples were collected in ECM monitors deployed at the same time as personal sampling; these monitors were not turned on. The samples correspond to visits conducted between August 2018 and February 2020, from baseline and five follow-up visits among participants assigned to the LPG stove group (intervention arm) and biomass stove group (control arm). Additional details on the exposure sampling visits during pregnancy are described by Johnson et al. [43].

Laboratory and data processing

The filter management procedures have been described elsewhere [43]. Briefly, filters were weighed on a 1-µg resolution MSA6.6s-000-DF Sartorius Cubis microbalance (Sartorius, Göttingen, Germany) to calculate $PM_{2.5}$ mass deposition and processed using the SootScan™ OT-21 optical transmissometer (Magee Scientific, Ljubljana, Slovenia) to obtain black carbon (BC) attenuations. All filter measurements were conducted at the Environmental Health Sciences Department at the University of Georgia (Athens, GA, USA). The equations and data corrections for estimating both $PM_{2.5}$ and BC concentrations are described in the Supporting Information (Eqs. S1 and S2).

The concentrations of the inorganic elements were quantified via energy dispersive X-ray fluorescence (ED-XRF) using an ARL QUANT'X

spectrometer (Thermo Scientific, MA, USA) at Colorado State University (Fort Collins, CO, USA). The elements resolved according to predefined quality standards were magnesium (Mg), aluminum (Al), silicon (Si), sulfur (S), potassium (K), calcium (Ca), titanium (Ti), chromium (Cr), manganese (Mn), iron (Fe), nickel (Ni), copper (Cu), zinc (Zn), gallium (Ga), arsenic (As), selenium (Se), cadmium (Cd), indium (In), tin (Sn), tellurium (Te), iodine (I), lead (Pb), chlorine (Cl) and sodium (Na), and their instrumental limits of detection (LoDs) were also provided. These concentrations were adjusted by subtracting the median blank filter concentration of every element, separately, and replacing all values below LoD with LoD/($2^{0.5}$) [47].

The analytical uncertainties of the concentrations for the inorganic elements were also recorded and provided by the ED-XRF instrument; these values were adjusted and combined using the error propagation method to determine total uncertainty (Supporting information: Eqs. S3–S5). The error propagation method calculates total uncertainty by summing the variance of every variable (assuming each one is independent of the others) [48]. Similar approaches were used to adjust the attenuation uncertainties (for BC estimates) and mass deposition uncertainties (for $PM_{2.5}$ estimates). Finally, the uncertainties below LoD were adjusted to be (5/6) x LoD [49].

Source apportionment

Source apportionment was conducted using the PMF receptor model from the U.S. Environmental Protection Agency (EPA PMF 5.0) [15]. The goal of PMF is to identify the factors (sources) and the amount of mass contributed by each factor to each individual sample in a mixture of pollutants. PMF solves the bilinear equation linking sources to observed concentrations

$$X_{ij} = \sum_{k}^{p} g_{ik} f_{kj} + e_{ij} \quad (1)$$

where X_{ij} is the concentration of species j on sample i, p is the number of factors (sources), f_{kj} is the concentration of species j in the source profile k, g_{ik} is the relative mass concentration of source k contributing to the sample i and e_{ij} is the residual for each species and sample. The PMF algorithm aims to provide the most appropriate solution using the non-negative constraints on the sources, by minimizing the objective function Q (Supporting information: Eq. S6).

PMF modeling was performed using a different number of factors ($p = 3$, 4, 5 and 6). The input compounds were BC, $PM_{2.5}$, and all inorganic elements with >50% of their observations above LoD (i.e., those having median values above their respective LoD). The observations were all the filter samples with non-missing concentrations. The criteria to select the most appropriate model for each selected number of sources was based on (1) the interpretation of sources based on profiles previously described in the literature; (2) the Q(true)/Q(expected) ratio, where Q(true) is the goodness-of-fit parameter calculated using all observations and the Q(expected) represents the observations without analytical difficulties or anomalous values; and (3) the displacement (DISP), bootstrap (BS) and BS-DISP errors, which estimate how well the PMF models distinguish the sources. Details about the error estimates are described elsewhere [15, 17, 18]. A comparison between observed and predicted $PM_{2.5}$ mass concentrations was also reported from the selected PMF model.

Contributions of species by study arm

The Wilcoxon signed rank test, a method that evaluates the significance of differences in experiments based on a ranking method [50], was computed to assess significant differences between the intervention and control groups with respect to their observed concentrations of $PM_{2.5}$, BC and inorganic elements. In addition, PMF output provides the average $PM_{2.5}$ mass concentration contributions categorized by study group, alongside their 95% confidence intervals. The study groups were divided into three categories: baseline, and two post-randomization groups (control and intervention). These calculations, alongside the data concentration estimates and uncertainty adjustments, were computed in R (version 4.3.2, R Foundation for Statistical Computing, Vienna, Austria).

Ethical approval

All methods were performed in accordance with the relevant guidelines and regulations. All participants provided written informed consent. The study protocol was reviewed and approved by institutional review boards (IRBs) or Ethics Committees at Emory University (00089799), Johns Hopkins University (00007403), Sri Ramachandra Institute of Higher Education and Research (IEC-N1/16/JUL/54/49) and the Indian Council of Medical Research—Health Ministry Screening Committee (5/8/4-30/(Env)/Indo-US/2016-NCD-I), Universidad del Valle de Guatemala (146-08-2016) and Guatemalan Ministry of Health National Ethics Committee (11-2016), Asociación Benefica PRISMA (CE2981.17), the London School of Hygiene and Tropical Medicine (11664-5) and the Rwandan National Ethics Committee (No.853/RNEC/2016), and Washington University in St. Louis (201611159). The study has been registered with ClinicalTrials.gov (Identifier NCT02944682).

RESULTS

Summary of samples

Six hundred and twenty-nine sample filters were available for estimating the concentrations of chemical species. One filter sample was excluded from analysis because it was damaged, and it was not possible to correctly estimate the species concentrations. A summary of the concentrations of all the species and their number of observations above LoD is given in Table 1. $PM_{2.5}$, BC, and 12 of the inorganic elements: Mg, Al, Si, S, K, Ca, Ti, Mn, Fe, Zn, Cl, and Na, were the chemical species used for PMF modeling. When categorizing the samples by study arm, 323 (51%) and 306 (49%) belonged to the control and intervention arm, respectively. Nine of the species' concentrations differed significantly between control and intervention arm samples, namely: Mg, S, K, Ca, Mn, Fe, Cl, BC and $PM_{2.5}$. These differences can be observed in Fig. 1, which also shows results from baseline (pre-treatment) samples, where results mirror those of the control group. Concentrations by study group are observed in more detail in the Supporting Information (Table S2).

Factor designation

An initial run of the PMF model helped determine which species had more weight than others. Based on the residuals, Mg, Al, and Ca were categorized as 'Weak' species; meanwhile, Zn was categorized as 'Bad' due to the low correlation between observed and predicted values ($R^2 < 0.2$).

According to the species profiles, the Q(true)/Q(exp) ratio, and the error estimates, the four-factor model was selected as the most robust to represent the number of sources for $PM_{2.5}$ in our personal exposure samples. A summary of the Q(true)/Q(exp) ratio and the error estimates for the models with 3 to 6 sources (factors) is shown in the Supporting information (Table S3 and Figs. S1–S2). Figure 2 shows the profiles of the four sources identified using PMF: biomass burning, crustal, other soil, and fossil fuels. It was observed that biomass burning was the highest source contributor to $PM_{2.5}$ (69.6%), followed by fossil fuels (22%), and then crustal and other soil in similar proportions (4.1% and 4.2%, respectively) (Fig. 3).

Source 1, the strongest contributor, was determined to be biomass burning due to the high proportion of K and Cl [38], which are indicators of combustion of vegetative material or smoke emitted as potassium salts [40, 51]. This source was expected, given the high proportion of biomass fuel burning in this location. Source 2 was determined to be crustal, which is characterized by Fe, Si, Ti, Mn and Ca in high proportions [22, 40]. This is indicative of typical soil composition, which, for this region of Guatemala, is rich in Fe. Source 3 was determined as another type of soil, which is identified by the presence of Mg, Al, and Na in high proportion, and the presence of Ca, Ti, Fe and BC [21, 26, 52]. We hypothesize that this is due to the presence of clay-made biomass stoves in most homes and potentially road dust [40]. This source, although in different proportions, has similar components as Source 2 (Fig. 3), which further supports this representing another type of soil [38], or a soil component from clay. Source 4 was determined to be a fossil fuel, characterized by the high proportion of S, which is likely associated with sulfur in vehicular fuel, which includes gasoline

Table 1. Summary of filter-based concentrations of all the detected chemical species ($n = 629$).

Species	Concentration (μg/m³)						N Samples Above LoD (%)
	Mean (SD)		Median (IQR)		Min	Max	
$PM_{2.5}$	93.0	(103)	55.0	(25.0–123.0)	9.7	780.0	571 (90.8)
BC	9.10	(7.10)	8.10	(3.40–13.00)	0.95	96.00	571 (90.8)
S	0.39	(0.28)	0.31	(0.20–0.48)	0.003	2.00	628 (99.8)
Ca	0.33	(0.29)	0.24	(0.15–0.42)	0.005	2.90	628 (99.8)
Mg	0.13	(0.09)	0.10	(0.07–0.16)	0.007	0.53	627 (99.7)
K	1.10	(1.10)	0.64	(0.34–1.3)	0.002	9.70	627 (99.7)
Fe	0.31	(0.27)	0.23	(0.12–0.42)	0.006	2.00	627 (99.7)
Cl	1.40	(2.30)	0.52	(0.15–1.7)	0.004	17.0	624 (99.2)
Al	0.67	(0.44)	0.57	(0.40–0.83)	0.008	2.60	599 (96.8)
Si	0.85	(0.86)	0.56	(0.19–1.20)	0.009	4.40	575 (91.4)
Na	0.14	(0.09)	0.12	(0.05–0.18)	0.052	0.52	538 (86.0)
Zn	0.012	(0.014)	0.005	(0.003–0.015)	0.003	0.180	487 (77.0)
Mn	0.019	(0.019)	0.015	(0.007–0.024)	0.004	0.150	472 (75.0)
Ti	0.040	(0.033)	0.030	(0.011–0.055)	0.011	0.280	465 (74.0)
Cu	0.016	(0.012)	0.010	(0.010–0.016)	0.010	0.080	280 (44.6)
I	1.20	(1.40)	0.41	(0.41–1.50)	0.41	9.50	270 (43.0)
Te	0.61	(0.66)	0.28	(0.28–0.69)	0.28	5.00	237 (37.8)
Ga	0.009	(0.007)	0.006	(0.006–0.009)	0.006	0.050	158 (25.2)
In	0.20	(0.14)	0.14	(0.14–0.14)	0.14	1.10	145 (23.2)
Sn	0.30	(0.18)	0.24	(0.24–0.24)	0.24	1.60	112 (17.9)
Ni	0.007	(0.004)	0.006	(0.006–0.006)	0.006	0.060	80 (12.9)
Cd	0.14	(0.06)	0.12	(0.12–0.12)	0.12	0.62	79 (12.7)
Cr	0.008	(0.005)	0.007	(0.007–0.007)	0.007	0.120	58 (9.4)
As	0.009	(0.004)	0.008	(0.008–0.008)	0.008	0.050	51 (8.3)
Pb	0.043	(0.030)	0.042	(0.042–0.042)	0.042	0.780	6 (1.1)
Se	0.023	(0.001)	0.023	(0.023–0.023)	0.023	0.040	1 (0.3)

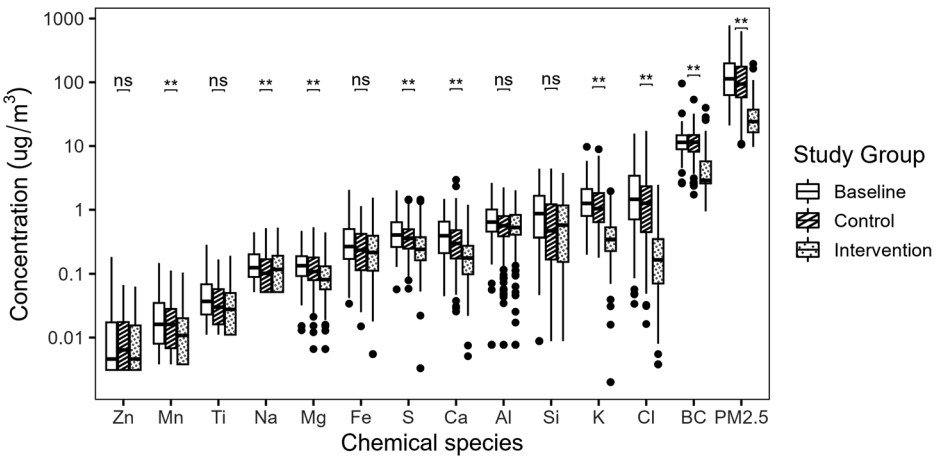


Fig. 1 Species concentrations selected for PMF modeling, categorized by study group. The figure shows the statistical comparison between Control and Intervention arm samples only, where **= statistically significant, ns = non-statistically significant, based on the Wilcoxon signed rank test.

and diesel. The proportion of BC and $PM_{2.5}$ [21, 53] (compared to the other species) also supports this source.

Predicted fine particulate matter ($PM_{2.5}$) contributions from the PMF model

In the PMF model, $PM_{2.5}$ was categorized as a weak contributor that does not affect the resulting solution since this species already includes the mass from other species (e.g., K, Mg, Cl, others) in the analysis. The predicted (reconstructed) $PM_{2.5}$ concentrations from the PMF model were plotted against the observed concentrations (Fig. 4). The final model was able to reproduce the mass concentrations and account for most of the variation ($\beta = 0.62 \pm 0.02$, $R^2 = 0.72$).

The highest predicted $PM_{2.5}$ concentrations were found in the baseline samples (Fig. 5) (mean: 155 μg/m³; 95% CI: 129.7, 180.2), followed by samples from post-intervention control (mean:

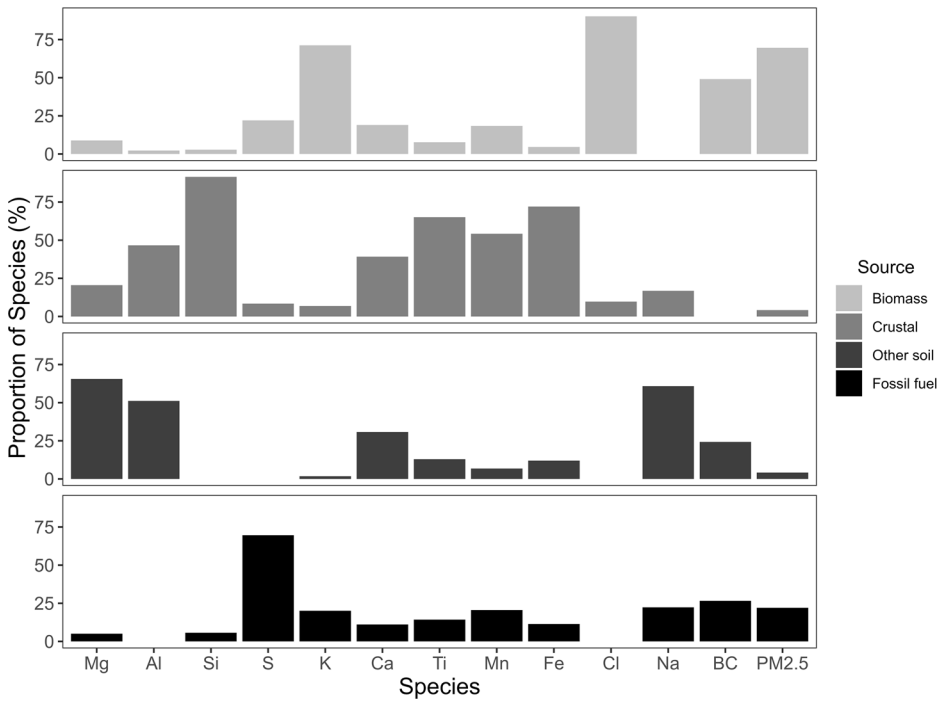


Fig. 2 Source profiles resolved from the four-factor model of $PM_{2.5}$. The chosen PMF model resolved four source profiles based on their chemical identities. These sources were identified as Biomass burning, Crustal, Other soil and Fossil fuel burning.

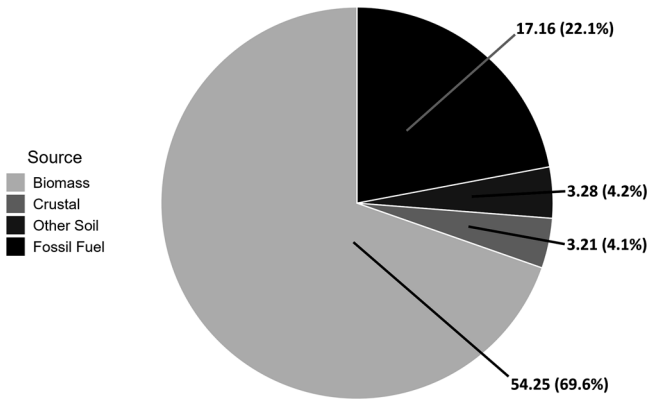


Fig. 3 $PM_{2.5}$ mass contribution by source. This chart shows the concentration (in μg/m^3) and percentage of contribution from each identified source for the total measured $PM_{2.5}$.

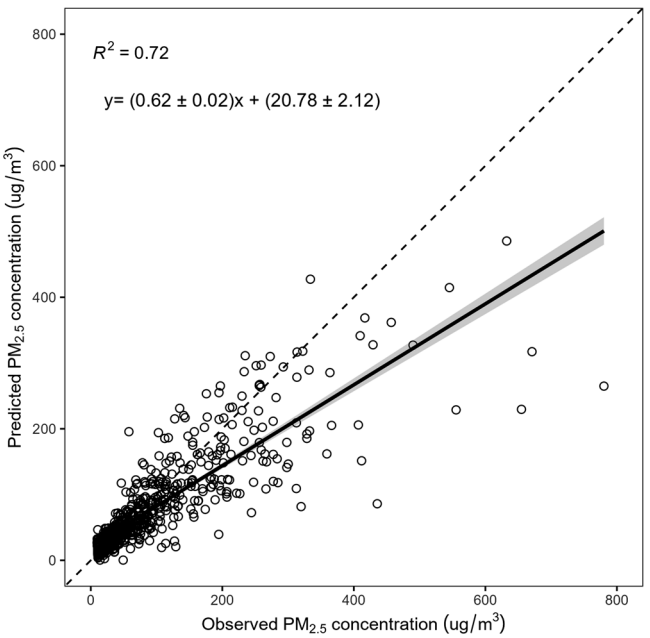


Fig. 4 Observed versus PMF model predicted $PM_{2.5}$ mass concentrations. The plot shows the observed vs predicted values (dots), showing the trend based on a linear regression model (solid line) alongside the confidence intervals (shaded area). The dashed line represents x = y.

127.2 μg/m^3; 95% CI: 114.7, 139.7), and the lowest predicted $PM_{2.5}$ concentrations were from the intervention samples (mean: 29.9 μg/m^3; 95% CI: 27.2, 32.7), post-intervention. If we look at all our samples, this represents 49.7% of mass contribution from baseline samples, followed by 40.7% for the control and 9.6% for the intervention, post-intervention samples. In the intervention samples, the concentrations are significantly lower compared to the post-intervention control samples, which are slightly below the low end of the CI as compared to baseline.

DISCUSSION

We identified four potential sources of $PM_{2.5}$ in rural Jalapa, Guatemala, based on personal exposure filter samples. The four sources (biomass burning, crustal, other soil and fossil fuels) were identified by PMF, one of the most used receptor models in source apportionment studies.

Figure 3 demonstrates that biomass burning and fossil fuels are the major sources driving $PM_{2.5}$ exposures in this area. This further supports the main idea from HAPIN that clean cooking interventions are needed to lower the concentrations of harmful pollutants, but as observed, it is also recommended to implement interventions that reduce exposures from other sources, like fossil fuel

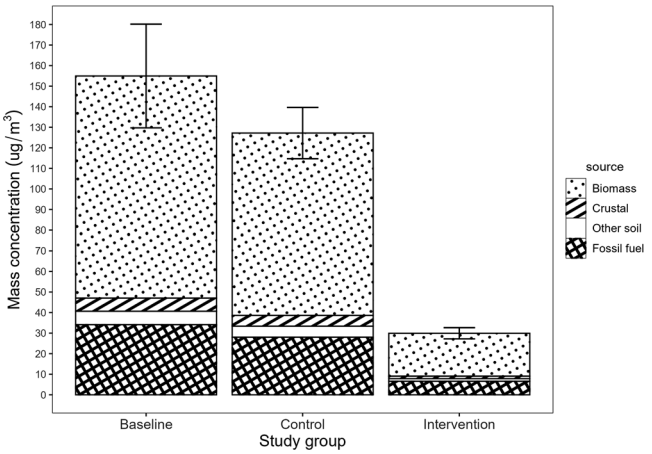


Fig. 5 Average contributions of $PM_{2.5}$ by source and study group. The figure shows the mean concentrations for each source identified, classified by the study group where the samples belonged to: Baseline, Control (post-intervention) and Intervention. The error bars are for the average $PM_{2.5}$ concentrations by study group.

emissions. The implementation of a clean cooking intervention decreased $PM_{2.5}$ concentrations and indirectly lowered the exposures from other sources (Fig. 5). The decrease of $PM_{2.5}$ in all sources in the intervention group can be explained by the reduced or eliminated use of biomass stoves inside the homes, and, since the analyzed samples belong to personal exposure measurements, there could have been behavioral changes associated with the intervention. One behavioral change could be a decrease in the time that the intervention arm participants spent outdoors, assuming that before receiving the intervention stove, they spent a significant amount of time outside collecting biomass fuel; therefore, being less exposed to vehicle emissions and other ambient particles associated with the crustal and other soil sources. The post-intervention control samples had lower $PM_{2.5}$ concentrations compared to baseline samples, which can be explained by a decrease in ambient $PM_{2.5}$ concentrations due to a lower use of biomass stoves in the intervention group households (assuming these are near each other). We acknowledge that the PMF model itself has associated uncertainties that play a role in the quantification of the identified and unidentified sources; this is observed when comparing the $PM_{2.5}$ concentrations by sources and by study group (Fig. 5). PMF can not necessarily distinguish between sources that tend to vary together, or two sources with very similar compositions [31, 53, 54].

In terms of PMF modeling, we chose the 4-factor model as the most robust. When comparing the predicted versus observed $PM_{2.5}$ concentrations (Fig. 4), this model showed high performance ($R^2 = 0.72$), although not as high as that observed in other source apportionment studies [21, 23, 55]. Our selection of the 4-factor model was also influenced by the Q(true)/Q(expected) ratio, the error estimates, and other parameters detailed in the Supporting information (Table S3 and Figs. S1–S3). Finally, the choice of species profiles was also driven by existing literature, which contributed to model choice, as well as identifying the potential sources of $PM_{2.5}$. As expected, biomass burning was identified as a predominant $PM_{2.5}$ source in these samples, given that they represent exposures from an area where a high proportion of the population relies on biomass for cooking. A HAPIN publication by Campbell et al. [56]. reported that wood represented 99% of the primary cooking fuel in the Guatemala site. This is further supported by the relatively high concentrations of potassium present in this factor compared to the others, which is a clear indicator of biomass burning.

The fossil fuel source was identified based on the high proportion of sulfur, which is an indicator of diesel rich in sulfur. The EPA began regulating the sulfur content in diesel in the US in 1993 [57], consistently decreasing it to ultra-low sulfur diesel (ULSD) at 15 ppm. In Guatemala, the regulations to decrease sulfur content began in 2016, establishing a 500 ppm level; but only 15 gas fuel stations, mostly located in urban areas, reported using ULSD [58, 59]. It is hypothesized that most of the fuel gas stations in Guatemala sold diesel with concentrations between 50 and 500 ppm during the time in which the study samples were collected, and there were no fuel stations that sold ULSD near the study location in Jalapa. The crustal and other soil sources were harder to differentiate from each other based on the PMF 4-factor model outputs (Supporting information: Table S3 and Fig. S3 "D"), but there are two reasons for their differences. It is hypothesized that there could be at least two types of soil that represent the micro-environments of this rural setting, and the model was able to differentiate between them. Also, it can be that the model was able to discriminate between particles emitted from the clay biomass stoves and the soil or floor material. However, these explanations would have to be confirmed by a more profound geological and chemical analysis of the types of soils and biomass stove materials.

When we considered which elements to include in the PMF model, the 50% above LoD cutoff point was selected, as a balance between maximizing the number of elements used and having those elements quantified in most samples. This process aids the model in more easily resolving the number of sources and avoids overestimating them based on elements with seasonal patterns. This also has a disadvantage, since this could have truncated the input from elements that are present in certain seasons, for example, crop burning seasons, holiday festivities or others. Inorganic elements that ultimately were not included in the PMF modeling due to LoD, mostly heavy metals, showed low concentrations compared to other XRF studies [60, 61]. Some of these could be associated with trash or plastic burning. A previous study by Kearns et al [62]. showed that 76% of participants from a subset of the HAPIN Guatemala study population preferred trash burning as their primary waste disposal method. A smaller proportion reported burning plastic, which could support the fact that some elements that are tracers for plastic or waste burning were observed in lower concentrations.

Our findings can be compared to previous source apportionment studies conducted in similar settings. Huang and collaborators [63] determined that wood combustion and cooking sources were the predominant sources of $PM_{2.5}$ in rural villages from China, accounting for around 77% of the $PM_{2.5}$ emissions in total. Zhou and colleagues [40] also determined that biomass was the predominant source of $PM_{2.5}$ in both of their study sites in Western Africa (39–62% and 74–87%). The second highest contributing source was identified as road dust and vehicle/mobile emissions in both studies, and the West Africa study mentions crustal as a minor source of $PM_{2.5}$ The Western Africa study relied on kitchen area and outdoor measurements instead of personal measurements; however the study population was more heterogeneous, for example the inclusion of different study areas and participants that did not used biomass fuels for cooking. On the other hand, the study from China collected personal exposure samples as we did, and the resolution of the model was improved due to the inclusion of more chemical species detected through additional chemical analyses. However, the sample population was smaller than ours, and it only included participants who were exposed to biomass fuel for cooking. Nevertheless, both study results are fairly consistent with our findings and show similarities between totally separate geographical locations but similar socioeconomic characteristics [63–65].

Although the results of this study are relevant, it is important to point out that certain chemical components were not analyzed and, therefore, not included for analysis. Organic carbon (OC) and elemental carbon (EC) represent an important fraction of the carbonaceous species [11], and they could be more representative than BC for total carbonaceous components in $PM_{2.5}$. Gaseous species such as ammonium (NH_4^+), sulfates and nitrates have been previously identified as tracers for secondary aerosols [32]. In addition, antimony (Sb) has been determined as an important marker of plastic burning [66]. Sb and other elements can be more easily incorporated in the analysis using ED-XRF, but OC and EC would require the use of the thermal reflectance method and the gaseous species by ion chromatography analysis. The last two methods are known to be destructive for the sample, so it would be advisable to perform this analysis at the end or with a different sampling filter. Another limitation of this study is that these results provide a snapshot of the chemical characterization of species, but the seasonal patterns are not determined. This would need an increased number of samples throughout the year and for multiple years to capture seasonal variation.

To our knowledge, this is the first source apportionment study conducted in Central America, and it is the first one reported from Guatemala. These results contribute to the growing body of literature for chemical characterization and will likely pave the road for future source apportionment studies in this region. The HAPIN study and similar studies are a step in the right direction for exposure reductions; however, as demonstrated by our results, source apportionment helps us see the remaining gaps in exposure assessment. This is proof that, besides clean cooking, other interventions that target hazardous emissions from other sources are needed. Although a risk assessment was not conducted, this study identifies and quantifies the sources of $PM_{2.5}$ so they can more easily be mediated by reducing the exposures through interventions from the public health perspective.

CONCLUSIONS

PMF modeling identified four potential sources of $PM_{2.5}$ in personal exposure samples from rural Jalapa, Guatemala. Biomass burning was identified as the source with the highest contribution to $PM_{2.5}$ mass (69.6%), followed by fossil fuel emissions (22%) and crustal and other soil sources in similar proportions (~ 4% each). When comparing sources by study group, baseline (50%) and control (41%) group samples experienced the largest amount of exposure to $PM_{2.5}$ mass, which further highlights biomass burning as the most prominent pollutant source in this location. The PMF model was able to accurately predict $PM_{2.5}$ mass concentrations ($R^2 = 0.72$) and account for the variation in the samples. This study demonstrates the potential to prioritize emissions reduction efforts in this location by implementing source-specific interventions, and it paves the road for future source apportionment studies.

DATA AVAILABILITY

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES

1. Gordon SB, Bruce NG, Grigg J, Hibberd PL, Kurmi OP, Lam KB, et al. Respiratory risks from household air pollution in low and middle-income countries. Lancet Respir Med. 2014;2:823–60. Oct.
2. IEA, IRENA, UNSD, World Bank, WHO. Tracking SDG7: The Energy Progress Report. Washington: World Bank; 2024.
3. Health Effects Institute. State of Global Air 2024. Special Report. Boston: Health Effects Institute; 2024.
4. Brauer M, Roth GA, Aravkin AY, Zheng P, Abate KH, Abate YH, et al. Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. The Lancet. 2024;403:2162–203.
5. Pope CA, Dockery DW. Health effects of fine particulate air pollution: lines that connect. J Air Waste Manag Assoc. 2006;56:709–42.
6. Younger A, Alkon A, Harknett K, Jean Louis R, Thompson LM. Adverse birth outcomes associated with household air pollution from unclean cooking fuels in low- and middle-income countries: a systematic review. Environ Res. 2022;204:112274.
7. Simkovich SM, Goodman D, Roa C, Crocker ME, Gianella GE, Kirenga BJ, et al. The health and social implications of household air pollution and respiratory diseases. Npj Prim Care Respir Med. 2019;29:1–17.
8. Naeher LP, Brauer M, Lipsett M, Zelikoff JT, Simpson CD, Koenig JQ, et al. Woodsmoke health effects: a review. Inhal Toxicol. 2007;19:67–106.
9. Thangavel P, Park D, Lee YC. Recent insights into particulate matter (PM2.5)-mediated toxicity in humans: an overview. Int J Environ Res Public Health. 2022;19:7511.
10. Viana M, Kuhlbusch TAJ, Querol X, Alastuey A, Harrison RM, Hopke PK, et al. Source apportionment of particulate matter in Europe: a review of methods and results. J Aerosol Sci. 2008;39:827–49.
11. Hopke PK, Dai Q, Li L, Feng Y. Global review of recent source apportionments for airborne particulate matter. Sci Total Environ. 2020;740:140091.
12. European Commission. Joint Research Centre. European guide on air pollution source apportionment with receptor models: revised version 2019. LU: Publications Office; 2019. Available from: https://data.europa.eu/doi/10.2760/439106
13. Gordon G, Pierson W, Daisey J, Lioy P, Cooper J, Watson JG, et al. Considerations for design of source apportionment studies. Atmos Environ. 1984;18:1567–82.
14. Reff A, Eberly SI, Bhave PV. Receptor modeling of ambient particulate matter data using positive matrix factorization: review of existing methods. J Air Waste Manag Assoc. 2007;57:146–54.
15. Norris G, Duvall R, Brown S, Bai S. Positive matrix factorization (PMF) 5.0 fundamentals and user guide.
16. Paatero P, Tapper U. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. Environmentrics. 1994;5:111–26.
17. Brown SG, Eberly S, Paatero P, Norris GA. Methods for estimating uncertainty in PMF solutions: examples with ambient air and water quality data and guidance on reporting PMF results. Sci Total Environ. 2015;518–519:626–35.
18. Paatero P, Eberly S, Brown SG, Norris GA. Methods for estimating uncertainty in factor analytic solutions. Atmospheric Meas Tech. 2014;7:781–97.
19. Jaeckels JM, Bae MS, Schauer JJ. Positive matrix factorization (PMF) analysis of molecular marker measurements to quantify the sources of organic aerosols. Environ Sci Technol. 2007;41:5763–9.
20. Martello DV, Pekney NJ, Anderson RR, Davidson CI, Hopke PK, Kim E, et al. Apportionment of ambient primary and secondary fine particulate matter at the Pittsburgh National Energy Laboratory particulate matter characterization site using positive matrix factorization and a potential source contributions function analysis. J Air Waste Manag Assoc. 2008;58:357–68.
21. Kim E, Hopke PK, Kenski DM, Koerber M. Sources of Fine Particles in a Rural Midwestern U.S. Area. Environ Sci Technol. 2005;39:4953–60.
22. Kim E, Hopke PK. Source apportionment of fine particles in Washington, DC, utilizing temperature-resolved carbon fractions. J Air Waste Manag Assoc. 2004;54:773–85.
23. Kim E, Hopke PK, Edgerton ES. Source identification of Atlanta aerosol by positive matrix factorization. J Air Waste Manag Assoc. 2003;53:731–9.
24. Brown SG, Lee T, Norris GA, Roberts PT, Collett JL, Paatero P, et al. Receptor modeling of near-roadway aerosol mass spectrometer data in Las Vegas, Nevada, with EPA PMF. Atmospheric Chem Phys. 2012;12:309–25.
25. Barrera V, Contreras C, Mugica-Alvarez V, Galindo G, Flores R, Miranda J. PM2.5 characterization and source apportionment using positive matrix factorization at San Luis Potosi City, Mexico, during the Years 2017–2018. Atmosphere. 2023;14:1160.
26. Mejía-Ponce LV, Hernández-López AE, Miranda-Martín-del-Campo J, Pineda-Santamaría JC, Reynoso-Cruces S, Mendoza-Flores JA, et al. Elemental analysis of PM10 in southwest Mexico City and source apportionment using positive matrix factorization. J Atmospheric Chem. 2022;79:167–98.
27. Martínez-Cinco M, Santos-Guzmán J, Mejía-Velázquez G. Source apportionment of PM $_{2.5}$ for supporting control strategies in the Monterrey Metropolitan Area, Mexico. J Air Waste Manag Assoc. 2016;66:631–42.
28. Dominutti PA, Mari X, Jaffrezo JL, Dinh VTN, Chifflet S, Guigue C, et al. Disentangling fine particles (PM2.5) composition in Hanoi, Vietnam: emission sources and oxidative potential. Sci Total Environ. 2024;923:171466.
29. Hsu CY, Chiang HC, Lin SL, Chen MJ, Lin TY, Chen YC. Elemental characterization and source apportionment of PM 10 and PM 2.5 in the western coastal area of central Taiwan. Sci Total Environ. 2016;541:1139–50.
30. Park J, Kim H, Kim Y, Heo J, Kim SW, Jeon K, et al. Source apportionment of PM2.5 in Seoul, South Korea and Beijing, China, using dispersion-normalized PMF. Sci Total Environ. 2022;833:155056.

31. Gadi R, Shivani, Sharma SK, Mandal TK. Source apportionment and health risk assessment of organic constituents in fine ambient aerosols (PM2.5): a complete year study over the National Capital Region of India. Chemosphere. 2019;221:583–96.
32. Sharma SK, Mandal TK, Jain S, Saraswati, Sharma A, Saxena M. Source apportionment of PM2.5 in Delhi, India using PMF model. Bull Environ Contam Toxicol. 2016;97:286–93.
33. Ramírez O, Sánchez de la Campa AM, Amato F, Catacolí RA, Rojas NY, de la Rosa J. Chemical composition and source apportionment of PM10 at an urban background site in a high–altitude Latin American megacity (Bogota, Colombia). Environ Pollut. 2018;233:142–55. Feb 1.
34. Carvalho JS, do Nascimento RKS, Cintra JVFDRF, da Rosa NLC, Grosseli GM, Fadini PS, et al. Source apportionment and health impact assessment of atmospheric particulate matter in the city of São Carlos, Brazil. Chemosphere. 2023;326:138450.
35. Pereira GM, Teinilä K, Custódio D, Gomes Santos A, Xian H, Hillamo R, et al. Particulate pollutants in the Brazilian city of São Paulo: 1-year investigation for the chemical composition and source apportionment. Atmospheric Chem Phys. 2017;17:11943–69.
36. Jorquera H, Barraza F. Source apportionment of ambient PM2.5 in Santiago, Chile: 1999 and 2004 results. Sci Total Environ. 2012;435–436:418–29.
37. Villalobos AM, Barraza F, Jorquera H, Schauer JJ. Chemical speciation and source apportionment of fine particulate matter in Santiago, Chile, 2013. Sci Total Environ. 2015;512–513:133–42.
38. Ofosu FG, Hopke PK, Aboh IJK, Bamford SA. Biomass burning contribution to ambient air particulate levels at Navrongo in the Savannah zone of Ghana. J Air Waste Manag Assoc. 2013;63:1036–45.
39. Pervez S, Dubey N, Watson JG, Chow J, Pervez Y. Impact of different household fuel use on source apportionment results of house-indoor RPM in central India. Aerosol Air Qual Res. 2012;12:49–60.
40. Zhou Z, Dionisio KL, Verissimo TG, Kerr AS, Coull B, Howie S, et al. Chemical characterization and source apportionment of household fine particulate matter in rural, Peri-urban, and Urban West Africa. Environ Sci Technol. 2014;48:1343–51.
41. Lai AM, Carter E, Shan M, Ni K, Clark S, Ezzati M, et al. Chemical composition and source apportionment of ambient, household, and personal exposures to PM2.5 in communities using biomass stoves in rural China. Sci Total Environ. 2019;646:309–19.
42. Clasen T, Checkley W, Peel JL, Balakrishnan K, McCracken JP, Rosa G, et al. Design and rationale of the HAPIN study: a multicountry randomized controlled trial to assess the effect of liquefied petroleum gas stove and continuous fuel distribution. Environ Health Perspect. 2020;128:047008.
43. Johnson MA, Steenland K, Piedrahita R, Clark ML, Pillarisetti A, Balakrishnan K, et al. Air pollutant exposure and stove use assessment methods for the household air pollution intervention network (HAPIN) trial. Environ Health Perspect. 2020;128:047009.
44. Barr DB, Puttaswamy N, Jaacks LM, Steenland K, Rajkumar S, Gupton S, et al. Design and rationale of the biomarker center of the household air pollution intervention network (HAPIN) trial. Environ Health Perspect. 2020;128:047010.
45. INE. XII Censo Nacional de Población y VII de Vivienda: Principales resultados del censo 2018. Guatemala: Instituto Nacional de Estadística de Guatemala; 2019. Available from: www.censopoblacion.gt
46. TessaDEM. Guatemala topographic map. 2023. Available from: https://en-us.topographic-map.com/map-fl23l/Guatemala/?center=14.61138%2C-90.11779&zoom=13
47. Hornung RW, Reed LD. Estimation of average concentration in the presence of nondetectable values. Appl Occup Environ Hyg. 1990;5:46–51.
48. Anderson GM. Error propagation by the Monte Carlo method in geochemical calculations. Geochim Cosmochim Acta. 1976;40:1533–8.
49. Polissar AV, Hopke PK, Paatero P, Malm WC, Sisler JF. Atmospheric aerosol over Alaska: 2. Elemental composition and sources. J Geophys Res Atmospheres. 1998;103:19045–57.
50. Wilcoxon F. Individual comparisons by ranking methods. Biol Bull. 1945;1:80–3.
51. Luo L, Zhang YY, Xiao HY, Xiao HW, Zheng NJ, Zhang ZY, et al. Spatial distributions and sources of inorganic chlorine in PM2.5 across China in Winter. Atmosphere. 2019;10:505.
52. Rai P, Furger M, El Haddad I, Kumar V, Wang L, Singh A, et al. Real-time measurement and source apportionment of elements in Delhi's atmosphere. Sci Total Environ. 2020;742:140332.
53. Gupta S, Gadi R, Sharma SK, Mandal TK. Characterization and source apportionment of organic compounds in PM10 using PCA and PMF at a traffic hotspot of Delhi. Sustain Cities Soc. 2018;39:52–67.
54. Srivastava D, Xu J, Vu TV, Liu D, Li L, Fu P, et al. Insight into $PM_{2.5}$ sources by applying positive matrix factorization (PMF) at urban and rural sites of Beijing. Atmospheric Chem Phys. 2021;21:14703–24.
55. Eatough DJ, Cropper P, Keeton W, Burrell E, Hansen JC, Farber R, et al. Apportionment of $PM_{2.5}$ adjacent to the I-710 Harbor Freeway in Long Beach, CA. J Air Waste Manag Assoc. 2020;70:260–82.
56. Campbell DA, Johnson M, Piedrahita R, Pillarisetti A, Waller LA, Kearns KA, et al. Factors determining black carbon exposures among pregnant women enrolled in the HAPIN trial. Environ Sci Technol. 2024;58:10162–74.
57. Environmental Protection Agency. Control of air pollution from new motor vehicles: Heavy-duty engine and vehicle standards and highway diesel fuel sulfur control requirements. 2001.
58. Bolaños R. Gasolineras distribuyen diésel con menos azufre. Prensa Libre. 2016. Available from: https://www.prensalibre.com/economia/se-cumple-norma-de-importacion/
59. Anonymous. Ya en Guatemala diésel bajo en azufre, ULSD. Perspectiva. 2016. Available from: https://www.perspectiva.gt/empresa/ya-guatemala-diesel-ultra-azufre-ulsd/
60. Zhang X, Eto Y, Aikawa M. Risk assessment and management of PM2.5-bound heavy metals in the urban area of Kitakyushu, Japan. Sci Total Environ. 2021;795:148748.
61. Shezi B, Jafta N, Naidoo RN. Potential health risks of indoor particulate matter and heavy metals in resource-constrained settings of South Africa. Atmosphere. 2024;15:911.
62. Kearns KA, Naeher LP, McCracken JP, Boyd Barr D, Saikawa E, Hengstermann M, et al. Estimating personal exposures to household air pollution and plastic garbage burning among adolescent girls in Jalapa, Guatemala. Chemosphere. 2024;348:140705.
63. Huang W, Baumgartner J, Zhang Y, Wang Y, Schauer JJ. Source apportionment of air pollution exposures of rural Chinese women cooking with biomass fuels. Atmos Environ. 2015;104:79–87.
64. Dionisio KL, Howie SRC, Dominici F, Fornace KM, Spengler JD, Donkor S, et al. The exposure of infants and children to carbon monoxide from biomass fuels in The Gambia: a measurement and modeling study. J Expo Sci Environ Epidemiol. 2012;22:173–81.
65. Johnson M, Pillarisetti A, Piedrahita R, Balakrishnan K, Peel JL, Steenland K, et al. Exposure contrasts of pregnant women during the Household Air Pollution Intervention Network randomized controlled trial. Environ Health Perspect. 2022;130:97005.
66. Turner A, Filella M. Field-portable XRF reveals the ubiquity of antimony in plastic consumer products. Sci Total Environ. 2017;584–585:982–9.

ACKNOWLEDGEMENTS

The investigators would like to thank the members of the advisory committee—Drs. Patrick Breysse, Donna Spiegelman, and Joel Kaufman—for their valuable insight and guidance throughout the implementation of the trial. We also wish to acknowledge all research staff and study participants for their dedication to and participation in this important trial. A multidisciplinary, independent Data and Safety Monitoring Board (DSMB) appointed by the National Heart, Lung, and Blood Institute (NHLBI) monitored the quality of the data and protected the safety of patients enrolled in the HAPIN trial. The DSMB consisted of: Catherine Karr (Chair), Nancy R. Cook, Stephen Hecht, Joseph Millum, Nalini Sathiakumar (deceased), Paul K. Whelton, Gail Weinmann and Thomas Croxton (Executive Secretaries). Program Coordination: Gail Rodgers, Bill & Melinda Gates Foundation; Claudia L. Thompson, National Institute of Environmental Health Sciences; Mark J. Parascandola, National Cancer Institute; Marion Koso-Thomas, Eunice Kennedy Shriver National Institute of Child Health and Human Development; Joshua P. Rosenthal, Fogarty International Center; Concepcion R. Nierras, NIH Office of Strategic Coordination—The Common Fund; Katherine Kavounis, Dong-Yun Kim, Barry S. Schmetter (deceased), and Antonello Punturieri, NHLBI. This research represents the NIH's contribution to the Global Alliance for Chronic Diseases (GACD) coordinated call for research on prevention and management of chronic lung diseases for 2016. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the US National Institutes of Health or Department of Health and Human Services.

AUTHOR CONTRIBUTIONS

Erick Mollinedo: Conceptualization, Software, Validation, Formal analysis, Investigation, Data curation, Writing—Original draft, Visualization, Supervision. Katherine A Kearns: Methodology, Conceptualization, Investigation, Writing—Review and editing. Armistead G Russell: Conceptualization, Writing—Review and editing, Supervision. Michael Johnson: Writing—Review and editing, Project administration. Christian L'Orange: Investigation, Resources, Writing—Review and editing. Ricardo Piedrahita: Methodology, Data curation, Formal analysis. Ajay Pillarisetti: Methodology, Data curation, Formal analysis, Writing—Review and editing. Jeremy A Sarnat: Conceptualization, Writing—Review and editing. John P McCracken: Supervision, Project administration. Lisa M Thompson: Writing—Review and editing, Supervision. Anaité

Diaz-Artiga: Writing—Review and editing, Supervision, Project administration. Maggie L Clark: Writing—Review and editing. Kyle Steenland: Writing—Review and editing. Lance A Waller: Writing—Review and editing, Visualization. Thomas F Clasen: Writing—Review and editing, Supervision, Project administration, Funding acquisition. Jennifer Peel: Project administration, Funding acquisition. William Checkley: Project administration, Funding acquisition. Luke P Naeher: Conceptualization, Resources, Writing—Review and editing, Supervision. HAPIN Investigators: Investigation.

FUNDING

The HAPIN trial was funded by the U.S. National Institutes of Health (cooperative agreement 1UM1HL134590) in collaboration with the Bill & Melinda Gates Foundation [OPP1131279].

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41370-026-00910-6.

Correspondence and requests for materials should be addressed to Luke P. Naeher.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

HAPIN INVESTIGATORS

Vigneswari Aravindalochanan[13], Kalpana Balakrishnan[13], Gloriose Bankundiye[14], Dana Boyd Barr[6], Vanessa Burrowes[15], Alejandra Bussalleu[16,17], Devan Campbell[1], Eduardo Canuz[8,17], Adly Castañaza[8], Howard H. Chang[11], Yunyun Chen[11], Marilú Chiang[17], Carmen Lucia Contreras[8], Rachel Craik[18], Victor G. Davila-Roman[19], Lisa de las Fuentes[19], Oscar de Leon[6,8], Priya D'Souza[6], Ephrem Dusabimana[14], Lisa Elon[11], Juan Gabriel Espinoza[17], Sarada S. Garg[13], Ahana Ghosh[2], Dina Goodman-Palmer[12,20], Savannah Gupton[6], Sarah Hamid[21], Stella M. Hartinger[16], Steven A. Harvey[15], Mayari Hengstermann[8], Ian Hennessee[6], Phabiola Herrera[12], Shakir Hossen[12], Marjorie Howard[11], Penelope P. Howards[21], Shirin Jabbarzadeh[11], Miles A. Kirby[22], Jacob Kremer[1], Margaret A. Laws[20], Grace Lee[6], Pattie Lenzen[19], Jiawen Liao[6], Amy E. Lovvorn[6], Jane Mbabazi[14], Eric D. McCollum[15], Julia N. McPeek[6], Rachel Meyers[19], J. Jaime Miranda[16], Libny Monroy[8], Lawrence Moulton[15], Alexie Mukeshimana[14], Krishnendu Mukhopadhyay[13], Bernard Mutariyani[14], Durairaj Natesan[13], Florien Ndagijimana[14], Laura Nicolaou[12], Azhar Nizam[11], Jean de Dieu Ntivuguruzwa[14], Parinya Panuwet[6], Aris T. Papageorghiou[18], Jennifer Peel[10], Irma Sayury Pineda Fuentes[8], Naveen Puttaswamy[13], Elisa Puzzolo[23,24], Sarah Rajkumar[10], Usha Ramakrishnan[25], Rengaraj Ramasami[13], Alexander Ramirez[8], Ghislaine Rosa[24], Joshua P. Rosenthal[26], P. Barry Ryan[6], Sudhakar Saidam[13], Sankar Sambandam[13], Saritha Sendhil[13], Suzanne Simkovich[27], Sheela S. Sinharoy[25], Kirk R. Smith[5], Gurusamy Thangavel[13], Ashley Toenjes[19], Lindsay J. Underhill[19], Viviane Valdes[28], Amit Verma[11], Jiantong Wang[11], Megan Warnock[11], Kendra N. Williams[12,15], Wenlu Ye[5], Bonnie N. Young[10] and Ashley Younger[18]

[13]Department of Environmental Health Engineering, Sri Ramachandra Institute of Higher Education and Research, Chennai, India. [14]Eagle Research Center, Kigali, Rwanda. [15]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA. [16]Universidad Peruana Cayetano Heredia, Lima, Peru. [17]Biomedical Research Unit, Asociación Benéfica PRISMA, San Miguel, Lima, Peru. [18]Nuffield Department of Women's and Reproductive Health, University of Oxford, Oxford, UK. [19]Washington University School of Medicine, Washington University in St. Louis, St. Louis, MO, USA. [20]Center for Non-Communicable Disease Research and Training, School of Medicine, Johns Hopkins University, Baltimore, MD, USA. [21]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA. [22]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [23]The Global LPG Partnership, New York, NY, USA. [24]Department of Public Health and Policy, University of Liverpool, Liverpool, UK. [25]Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA. [26]Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA. [27]Division of Healthcare Delivery Research, MedStar Health Research Institute, Columbia, MD, USA. [28]Division of Developmental Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA.