Supporting Information

Zhou et al. 10.1073/pnas.1019183108

SI Text

Pollutant Measurement and Analytical Methods. The cooking-area monitors were placed about 1-m above ground level and 1-m away from cooking stoves, and the living-area monitors were installed along the wall. Because we had a limited number of integrated monitors and because operating and maintaining monitors in distant neighborhoods was time-intensive, we restricted simultaneous operation to two households in the same neighborhood per 48 h. Measurements took place between November 2006 and August 2007, with the measurement schedule for each neighborhood provided elsewhere (1).

Blanks and duplicates (i.e., side-by-side measurements) were collected in multiple homes. We collected a total of 113 $PM_{2.5}$ and 113 PM_{10} samples (Table S2). These samples included 14 $PM_{2.5}$ duplicates, 13 PM_{10} duplicates, 9 $PM_{2.5}$ blanks, and 10 PM_{10} blanks. In addition, we conducted repeated 48-h $PM_{2.5}$ and PM_{10} measurements in four and five households, respectively. We calculated coarse PM ($PM_{2.5–10}$) as the difference between measured PM_{10} and $PM_{2.5}$ for each household or rooftop site.

Integrated PM. Methods for measurements at rooftop ambient sites have been described in a previous article (1). All integrated PM household samples were collected on PTFE filters with a ring (Pall Life Sciences; Teflo, 0.2-μm pore size, 37-mm diameter), back-supported by a Whatman drain disk. PM_{10} measurements used a personal exposure monitor (PEM) with a D_{50} of 10 μm (aerodynamic diameter) at 4 L per minute (LPM) (±10%), with an internal level greased well serving as the impaction surface. Most $PM_{2.5}$ measurements used a Harvard impactor (2, 3) with two size-selective inlets in series, each with a preoiled impactor plate serving as the impaction surface to reduce the effects of particle bounce. Each size-selective inlet had a D_{50} of 2.5 μm at 4 LPM (±10%). A few $PM_{2.5}$ measurements used a PEM with a D_{50} of 2.5 μm at 4 LPM (±10%) and an internal level greased well serving as the impaction surface. Airflow rates through the filters were measured at the start and end of each 48-h measurement period using a calibrated rotameter. Equipment and flow rates were checked at 24 h and adjusted as needed.

All filters were weighed before and after measurement on a Mettler Toledo MT5 microbalance at the Harvard School of Public Health (HSPH) Laboratory, after being conditioned in a temperature and humidity controlled environment [20.5 ± 0.2 °C, 39 ± 2% relative humidity (RH)] for at least 24 h and statically discharged via a polonium source. In both pre- and postweighing, filters were weighed twice; if these two masses were not within 5 μg of one another, they were weighed a third time. The mean of the two masses within 5 μg was used for calculating concentrations. After every batch of 10 filters, the zero, span, and linearity of the balance were checked via a set of class "S" weights. Final filter weights were adjusted using an air buoyancy correction (4).

Measured concentrations were used only if the pumps operated for ≥ 85% of the 48-h measurement period and if the average flow rate was within 15% of the intended rate. These criteria excluded 10 $PM_{2.5}$ and 11 PM_{10} integrated measurements. One additional PM_{10} measurement was excluded because of a broken connection in the airflow system. All PM concentrations were blank-corrected. The mean weights of blank samples were 2.2 and 4.4 μg, respectively, for $PM_{2.5}$ and PM_{10}. All filters weighed well above the limit of detection (calculated as three times the SD of the blanks), with the lowest filter weight being over 18- and 23-times larger than the limit of detection for $PM_{2.5}$ and PM_{10}, respectively.

Where duplicate measurements were taken, the average of the two measurements was used for analysis. The mean relative percent-difference of $PM_{2.5}$ and PM_{10} duplicate measurements were 7% and 5% respectively, each excluding one duplicate measurement with low flow rate. These values are consistent with other studies at the Environmental Protection Agency Center for Ambient Particle Health Effects at HSPH.

Continuous PM. Methods for continuous PM measurements at rooftop ambient sites are described in detail elsewhere (1). We measured continuous $PM_{2.5}$ using SidePak Model AM510 and DustTrak Model 8520 monitors (TSI Inc.) in household cooking and living areas, respectively. Both monitors have a built-in laser photometer that uses a 90° light scattering laser diode to measure airborne $PM_{2.5}$ concentration. SidePaks and DustTraks were operated at 0.8 LPM and used an upstream single or double mini-PEM as the external size-selective inlet, with an internal level greased well serving as the impaction surface. Mini-PEMs were changed at maximum intervals of 48 h, and more frequently if necessary. All DustTraks and SidePaks were calibrated daily to a zero filter. DustTrak and SidePak monitors measure PM every second. Data were recorded at 1-min intervals, with each record corresponding to the average of the previous 60 measurements. The internal time maintained by the monitors was periodically synchronized with an external time source. To ensure data validity, we excluded continuous data when the instrument malfunctioned (e.g., laser failure) or when flow rate was low because of low battery or a folded tube between the inlet and the instrument.

PM concentrations measured using light scattering are subject to error, because factory calibrations use specific aerosols whose characteristics (e.g., shape, size, density, and refractive index) may differ from those in field studies, and because factors such as RH affect measurements (5–7). We adjusted continuous PM in a three-step process. In the first step, we standardized the minute-by-minute records for the effects of RH, using relationships from previous studies (5). In the second step, we corrected all minute-by-minute PM records in each 48-h measurement period using a correction factor (CF) so that the average of RH-standardized continuous PM was equal to the integrated gravimetric PM level over the same period and at the same location. In the final step, we used a nonparametric regression (locally weighted scatterplot smoothing), with a 60-min bounding radius, for smoothing the continuous PM. In the above approach, the first step removes the effect of RH variation on measured PM within a single day; the second step ensures that the measurements are corrected against the gravimetric measurement, which has substantially less error than nephelometers; and the third step eliminates perturbations sustained for < 10 min but maintains longer-lasting patterns (8). We used smoothing because it may be possible that nephelometer measurement error is systematically different at higher concentrations. The median (interquartile range) of CFs was 1.00 (0.84–1.19).

We calculated CF directly for continuous $PM_{2.5}$ in the cooking areas, where integrated $PM_{2.5}$ was also measured. Neighborhood geometric mean of all cooking area CFs was used for cooking areas where the integrated sample was excluded for the above reasons or when the duration of cooking area continuous data were < 85% of the 48-h measurement period. We applied the cooking area CFs for continuous $PM_{2.5}$ in the living area of the same household.

Household and Neighborhood Socioeconomic Status, Fuel, and Housing. We used a structured questionnaire to collect data on number of household members, housing and cooking area characteristics, ownership of assets, fuels and stoves used for domestic and small-commercial cooking, and the presence of smokers in the house and other combustion sources. We used a Garmin Etrex GPS unit (Garmin Limited) to record geographic coordinates for each household, which were used to calculate the average distance between the household and main roads in ArcMap 9.3 (ESRI Corp.).

Following previous analyses of household data in developing countries (9, 10), we measured each household's socioeconomic status (SES) using an index based on housing characteristics, water and waste systems, and ownership of durable assets. To calculate the SES index based on empirical weights for each asset, we used principal component analysis of the set of assets recorded in the questionnaire. Principal component analysis is a statistical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables, the principal components (11). We used the first principal component as the SES index because it is a linear combination of all of the individual asset variables that explains the maximum variance across the households (29% of total variance). The variables used in calculating the SES index were wall and floor materials, toilet and bathing facilities, solid and liquid waste-disposal methods, water source, and ownership of a sewing machine, telephone, cell phone, refrigerator, radio, television, electric iron, bicycle, motorcycle, or car.

We used a 10% sample of the Ghana 2000 Population and Housing Census to calculate the proportion of households who used biomass fuels in the Census Enumeration Area (EA) in which each study household was located. We also used the census data to calculate a community SES index for the household's EA. The EA SES used data on type and tenure of dwelling, materials of outer walls, floor, and roof, toilet and bathing facilities, solid and liquid waste-disposal methods, water source, and the number of persons per room and per bedroom; data on all these variables were available in the Ghana 2000 Population and Housing Census. The PCA was conducted using data from individual census records (each corresponding to one household), with household scores averaged to obtain EA SES.

Meteorological and Weather Variables. We obtained hourly RH and weather data from a station near the Accra International Airport, maintained by the National Oceanic and Atmospheric Administration, United States Department of Commerce. For hours with missing RH data, we predicted RH using simple linear models when data were missing for less than 3 h. When more than 3 h of data were missing, we used the average of RH for the same hour over 5 d before and 5 d after the missing value. We fitted a cubic spline function to hourly RH to obtain RH for each minute. We calculated the number of hours with rain in each 48-h measurement period using variables on present weather and past weather, which report precipitation for the current (t) and previous hour ($t - 1$). If these two variables had missing values, we used data on precipitation during the preceding 6 h to assign precipitation status.

Statistical Analysis. In addition to descriptive statistics, we used regression analysis to examine the association of cooking area PM with its potential household and neighborhood determinants that may be proxies for PM sources and for ventilation. We applied the following regression equation:

$$\begin{aligned} Ln(PM_{cooking\ area}) = \beta_0 &+ \beta_1 \times household\ fuel(own) \\ &+ \beta_2 \times household\ fuel(small-commercial) \\ &+ \beta_3 \times EA\ biomass\ use \\ &+ \beta_4 \times cooking\ location \\ &+ \beta_5 \times household\ size + \beta_6 \times SHS \\ &+ \beta_7 \times distance\ to\ main\ roads \\ &+ \beta_8 \times precipitation(+\beta_9 \times Ln(PM_{ambient})) \\ &+ \varepsilon \end{aligned}$$

where,

$PM_{cooking\ area}$: 48-h integrated PM in the cooking area;
Household fuel (own): biomass, nonbiomass;
Household fuel (small-commercial): none, biomass, nonbiomass;
EA biomass use: percentage of households using biomass in the household's EA;
Cooking location: inside the house, open air, separate cookhouse;
Household size: number of household members;
SHS (secondhand smoke): smokers in the house, no smokers in the house;
Distance to main roads: weighted average of distances from the household to main roads;
Precipitation: number of hours with precipitation during the 48-h measurement period; and
$PM_{ambient}$: average of 48-h integrated PM at nontraffic rooftop sites of the household's neighborhood in the same measurement period as that of the household.

We repeated our analysis with (model 1) and without (model 2) $PM_{ambient}$ in the regression to reflect the possibilities that neighborhood ambient PM may influence household concentrations (hence the need to adjust for $PM_{ambient}$) or that in a primarily biomass-using neighborhood, ambient PM may itself be largely because of household fuel use (hence a potential overadjustment).

Ventilation parameters other than cooking location were not included in the model because houses in the study neighborhoods were naturally ventilated, allowing particles to readily penetrate in and out the house. PM concentrations were log-transformed to ensure that the residuals were normally distributed. All analyses were conduct using the open-source statistical analysis package R version 2.8.

Comparison with Previous Studies. A study in Dhaka, Bangladesh (12) reported that higher-income households had lower kitchen PM_{10} concentrations, which is consistent with our findings, and with patterns of fuel use in our study households and neighborhoods. Another study in Accra found higher personal PM exposure among wood users, who were generally from lower-income households (13). Prior studies in Delhi, India (14) and Dhaka, Bangladesh (12) also found that clean fuels were associated with lower household PM levels. In Agra, India indoor-to-ambient ratios for $PM_{2.5}$ ranged from 0.76 to 1.13 (15), a smaller but consistent range compared with the 0.44 to 2.87 range in our study. The differences in the ranges could be because our study households and neighborhoods, which covered a large range of SES, had larger variation in fuel use, building design and material, and ventilation.

Strength and Limitation. Our study has a number of innovations and strengths: The study combined geo-referenced data from the census, household questionnaire, Accra road map, and household and ambient PM measurements to have uniquely complete data on pollution and its potential determinants at the household as well as community levels. The data were from four neighborhoods that covered the full range of community SES in Accra. Our continuous data allowed analyzing the temporal patterns of PM.

Furthermore, we used integrated PM data to correct for the measurement error of continuous data measured using light scattering.

The data used in this study also have a number of limitations. First, it would have been ideal to conduct measurements in a larger number of households, but this was beyond our resources. Second, data were collected in different months in the four neighborhoods. Adjustment for ambient PM in the multivariate analysis helped overcome macrolevel seasonal differences across the neighborhoods but it would be ideal to have multiple measurement campaigns in each neighborhood, in different seasons. Because of lack of data from different seasons, our results should not be used to estimate the usual or average pollution in these households. Third, although continuous data helped analyze the temporal patterns of pollution, PM measured using light scattering is subject to error. Although we systematically applied a CF to PM data, the steps involved in calculating CFs introduce additional uncertainty, especially for the living area continuous PM data, which was corrected using the cooking area CF. These strengths and limitations should inform the design of future research on the determinants of household pollution in developing country cities.

1. Dionisio KL, et al. (2010) Air pollution in Accra neighborhoods: Spatial, socioeconomic, and temporal patterns. *Environ Sci Technol* 44:2270–2276.
2. Marple VA, Willeke K (1976) Impactors design. *Atmos Environ* 10:891–896.
3. Marple VA, Rubow KL, Turner W, Spengler JD (1987) Low flow rate sharp cut impactors for indoor air sampling: Design and calibration. *JAPCA* 37:1303–1307.
4. Schoonover RM, Jones FE (1981) Air buoyancy correction in high-accuracy weighing on analytical balances. *Anal Chem* 53:900–902.
5. Chakrabarti B, Fine PM, Delfino R, Sioutas C (2004) Performance evaluation of the active-flow personal DataRAM $PM_{2.5}$ mass monitor (Thermo Anderson pDR-1200) designed for continuous personal exposure measurements. *Atmos Environ* 38: 3329–3340.
6. Day DE, Malm WC, Kreidenweis SM (2000) Aerosol light scattering measurements as a function of relative humidity. *J Air Waste Manag Assoc* 50:710–716.
7. Laulainen NS (1993) *Summary of Conclusions and Recommendations from a Visibility Science Workshop: Technical Basis and Issues for a National Assessment of Visibility Impairment (Pacific Northwest Laboratory PNL-8606), US Department of Energy* (Department of Energy, Washington, DC).
8. Cleveland WS, Grosse E, Shyu WM (1992) *Statistical Models in S*, eds Chambers JM, Hastie TJ (Wadsworth, Pacific Grove, CA), pp 309–376.
9. Wagstaff A (2000) Socioeconomic inequalities in child mortality: Comparisons across nine developing countries. *Bull World Health Organ* 78:19–29.
10. Gakidou E, et al. (2007) Improving child survival through environmental and nutritional interventions: The importance of targeting interventions toward the poor. *JAMA* 298:1876–1887.
11. Jolliffe IT (2002) *Principal Component Analysis* (Springer, NY).
12. Dasgupta S, Huq M, Khaliquzzaman M, Pandey K, Wheeler D (2006) Indoor air quality for poor families: New evidence from Bangladesh. *Indoor Air* 16:426–444.
13. Songsore J, McGranahan G (1998) The political economy of household environmental management: Gender, environment and epidemiology. *World Dev* 26:395–412.
14. Saksena S, et al. (2003) Exposure of infants to outdoor and indoor air pollution in low-income urban areas—A case study of Delhi. *J Expo Anal Environ Epidemiol* 13:219–230.
15. Massey D, Masih J, Kulshrestha A, Habil M, Taneja A (2009) Indoor/outdoor relationship of fine particles less than 2.5 μm ($PM_{2.5}$) in residential homes locations in central Indian region. *Build Environ* 44:2037–2045.

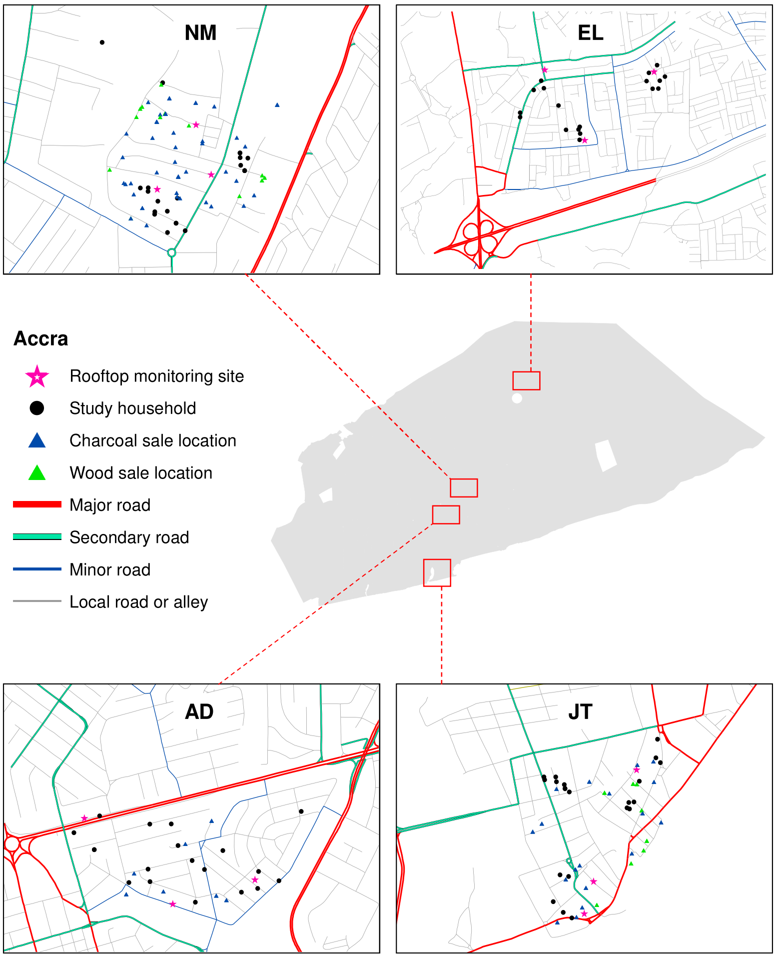

Fig. S1. Study neighborhoods, ambient measurement sites, and study households. We recorded biomass fuel sales locations through a neighborhood-wide census.

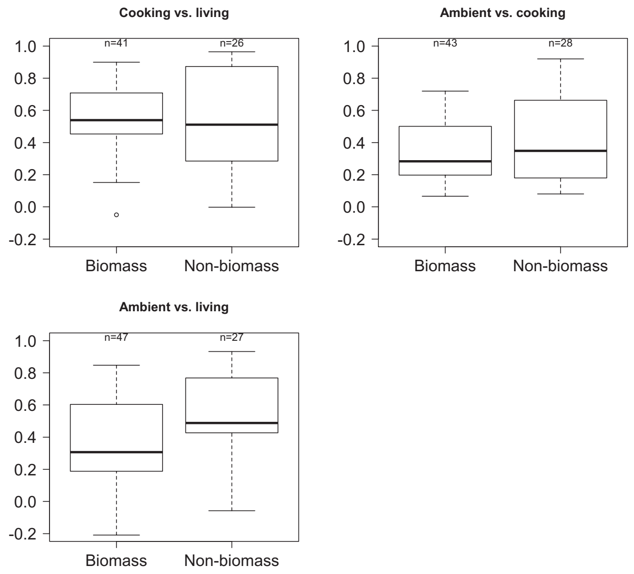

Fig. S2. Correlation coefficients of continuous $PM_{2.5}$ measurements in ambient, cooking area, and living area, stratified by household fuel use. Each correlation coefficient is calculated between minute-by-minute measurements over a 48-h measurement period for one household. The number of households in each group is *n*. In each plot, the middle line shows the median, and the bottom and top of the box show the 25th and 75th percentiles of data.

Table S1. Summary statistics for fuel, housing, and other characteristics of study households, by neighborhood

	JT	AD	NM	EL
Number of households in the analysis	21	20	21	18
Average number of residents per household	8.6	4.4	6.6	6.7
Average distance to main roads (m)	42	78	100	250
Percent of all households (%)				
Uses biomass fuel for household's own cooking	95	50	86	22
Fuel for small-commercial cooking				
No small-commercial cooking	55	75	71	83
Biomass	45	15	29	6
Other	0	10	0	11
Dwelling				
Compound room	90	65	100	6
Separate house	10	30	0	89
Other	0	5	0	6
Cooking Area				
Inside the house	10	40	10	83
Separate cookhouse	14	15	19	0
Open air	76	45	71	17
Source of drinking water				
Public tap	33	5	38	0
Piped into yard/plot	29	5	29	6
Sachet water	19	60	5	22
Piped into dwelling	19	25	29	33
Bottled water	0	5	0	39
Type of toilet				
Bucket/pan	10	10	10	0
Traditional toilet	86	10	67	11
Ventilated improved pit latrine	5	0	14	0
Flush toilet	0	80	10	89
Bathing facility				
Private open cubicle	0	0	0	6
Shared bathroom	67	50	48	17
Own bathroom	33	50	52	78
Solid waste disposal method				
Solid waste is collected	19	100	19	61
Burnt, public or other dump, buried, and other	81	0	90	39
Liquid waste disposal method				
Sewage system	0	0	0	61
Thrown on street, gutter, or compound, and other	100	100	100	39
Presence of smokers in the household	0	0	10	10

Table S2. Details of PM data collection by neighborhood

Neighborhood	Number of samples
JT	
48-h integrated $PM_{2.5}$ measurements in the cooking area*	20
48-h integrated PM_{10} measurements in the cooking area*	20
Household-days of continuous $PM_{2.5}$ data in the cooking area	25
Household-days of continuous $PM_{2.5}$ data in the living area	38
AD	
48-h integrated $PM_{2.5}$ measurements in the cooking area*	19
48-h integrated PM_{10} measurements in the cooking area*	19
Household-days of continuous $PM_{2.5}$ data in the cooking area	35
Household-days of continuous $PM_{2.5}$ data in the living area	35
NM	
48-h integrated $PM_{2.5}$ measurements in the cooking area*	25
48-h integrated PM_{10} measurements in the cooking area*	26
Household-days of continuous $PM_{2.5}$ data in the cooking area	53
Household-days of continuous $PM_{2.5}$ data in the living area	37
EL	
48-h integrated $PM_{2.5}$ measurements in the cooking area*	16
48-h integrated PM_{10} measurements in the cooking area*	14
Household-days of continuous $PM_{2.5}$ data in the cooking area	33
Household-days of continuous $PM_{2.5}$ data in the living area	34

*Number of valid samples, excluding duplicates.