Supporting Information for

Biomass fires increase infant mortality globally

Hemant K Pullabhotla, Mustafa Zahid, Sam Heft-Neal, Vaibhav Rathi, Marshall Burke

Corresponding Author Hemant K Pullabhotla
E-mail: h.pullabhotla@deakin.edu.au

This PDF file includes:

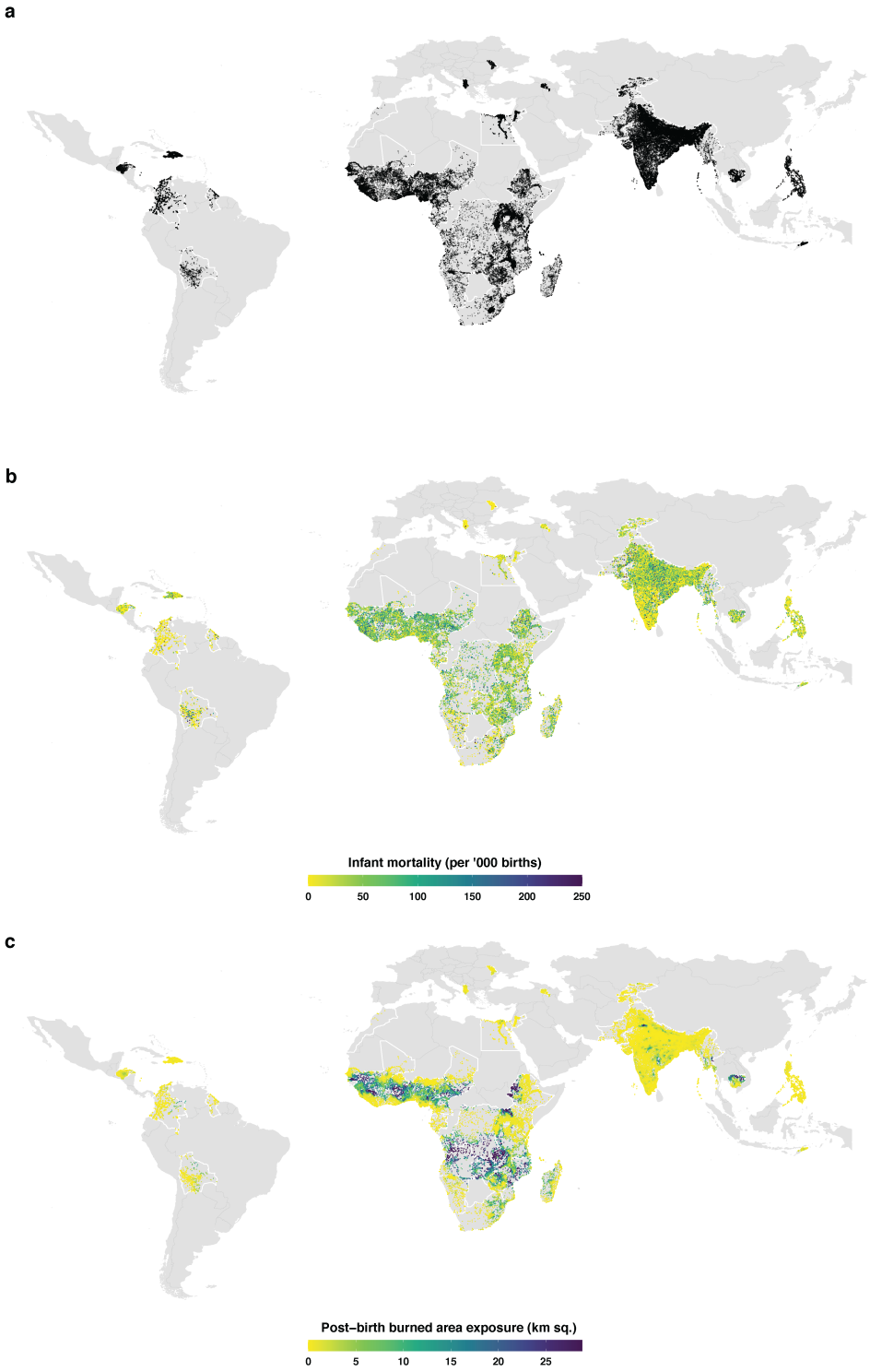

Fig. S1. Location of births in estimation sample, average infant mortality and burned area exposure. **a** Location of sample clusters in the DHS data used for estimation ($N = 93063$). DHS data provides geographic coordinates for households at the sample cluster level. **b** Cluster-level sample average infant mortality rate (deaths per '000 births) over the estimation period 2004 - 2018. **c** Cluster-level average post-birth exposure to biomass burned area in (square kilometers per month). Exposure is based on monthly burned area recorded in the up-wind quadrant within a 30 km distance, in the 12 months after birth. The range of values in **b** and **c** are capped at the 99^{th} percentile of the distribution for better visualization. White borders highlight the countries in the DHS births sample used in the estimation.

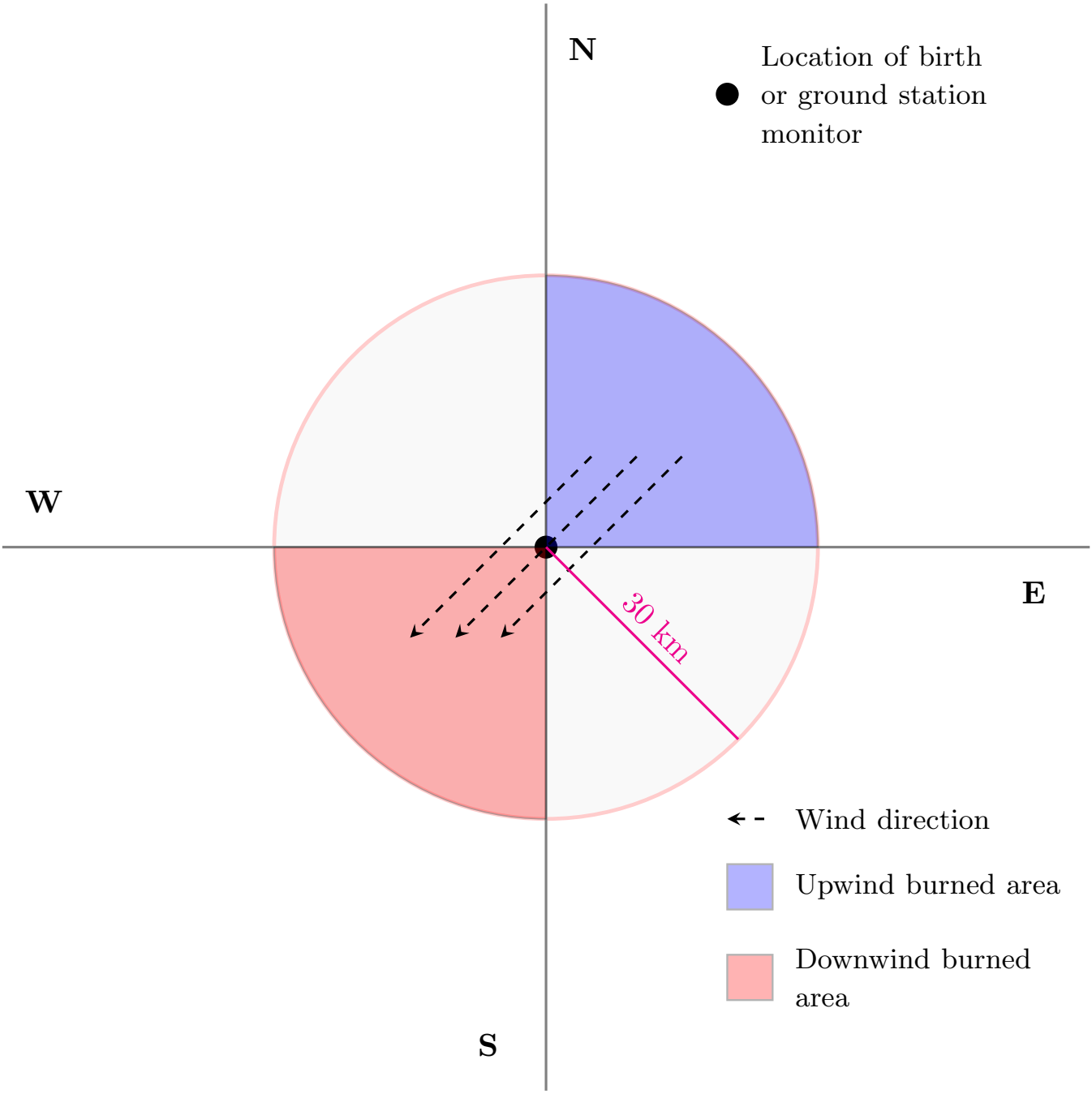

Fig. S2. Schematic showing definition of up and downwind burned areas. We estimate exposure to up and downwind biomass burned areas for each birth based on the prevailing wind direction (based on climate reanalysis data) for each month in the pre- and post-birth period. Exposure is calculated by taking the average monthly up and downwind burned areas for the pre- and post-birth periods. In the example shown here, the wind is blowing from the northeast to the southwest. Therefore, upwind (relative to the birth location or ground station monitor) burned area is the biomass burned in the northeast quadrant within a 30-km radius. Biomass burning in the opposing quadrant forms the downwind burned area. The same procedure is used to define up and downwind burned areas around air pollution ground monitors at a monthly resolution. We hypothesize that upwind burned area increases air pollution and result in adverse health outcomes.

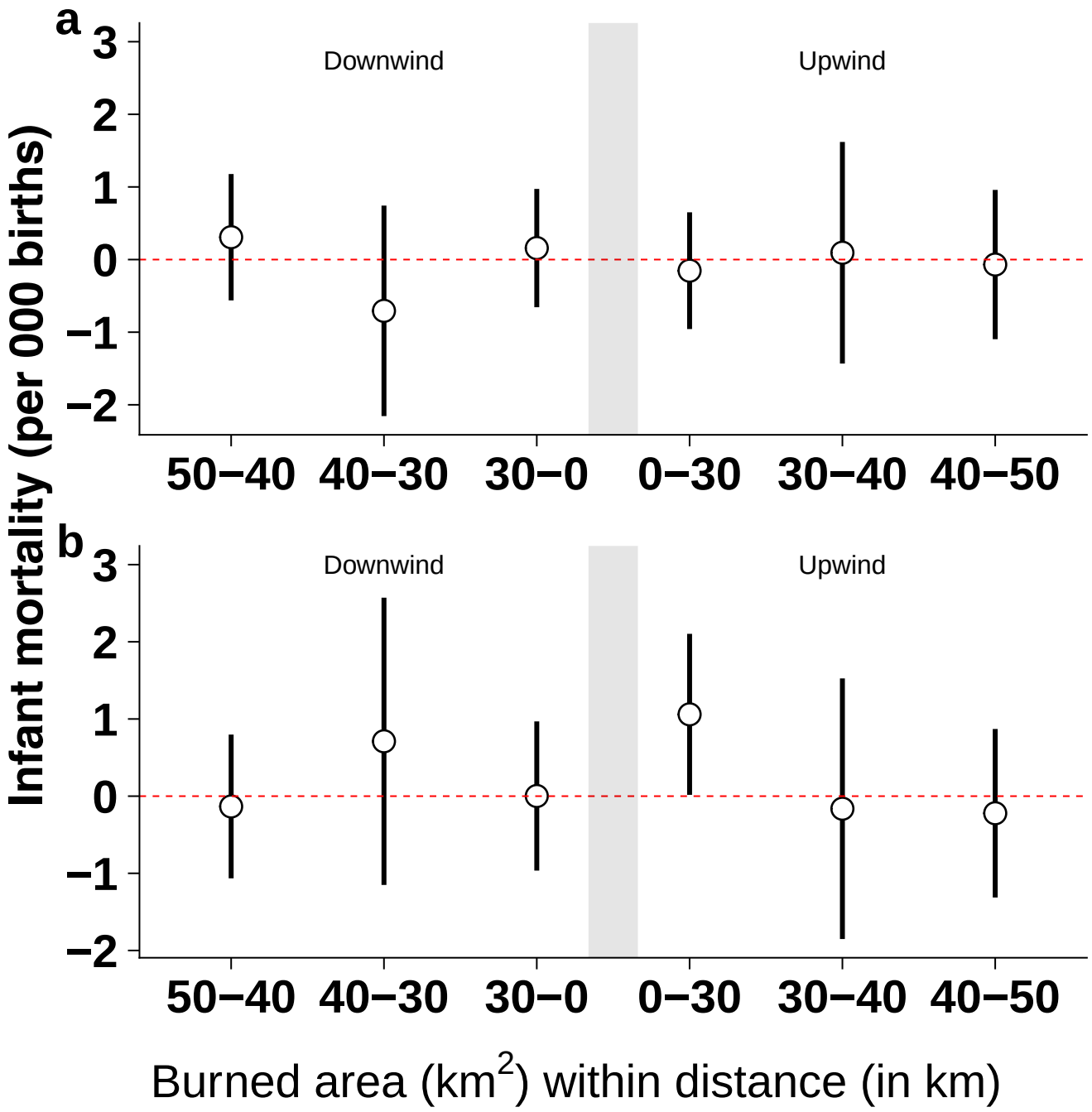

Fig. S3. Regression estimates of pre- and post-birth exposure to outdoor biomass burning on infant mortality. **a** and **b** show the coefficient estimates from the regression model in Equation [1] (Methods) for pre-birth and post-birth exposure, respectively. Circles indicate point estimates, and whiskers the 95% confidence interval on the point estimate. **a** Pre-birth exposure does not have an effect on risk of infant mortality. **b** Post-birth exposure to nearby biomass burning (within 30 km) in the upwind direction increases infant mortality by 1.06 deaths for an additional 1 km^2 increase in burned area. Biomass burning that is further away (30 to 40 or 40 to 50 km) has no effect. Downwind burned areas do not have a significant effect on infant mortality – consistent with pollution from burning blowing away from the location of birth.

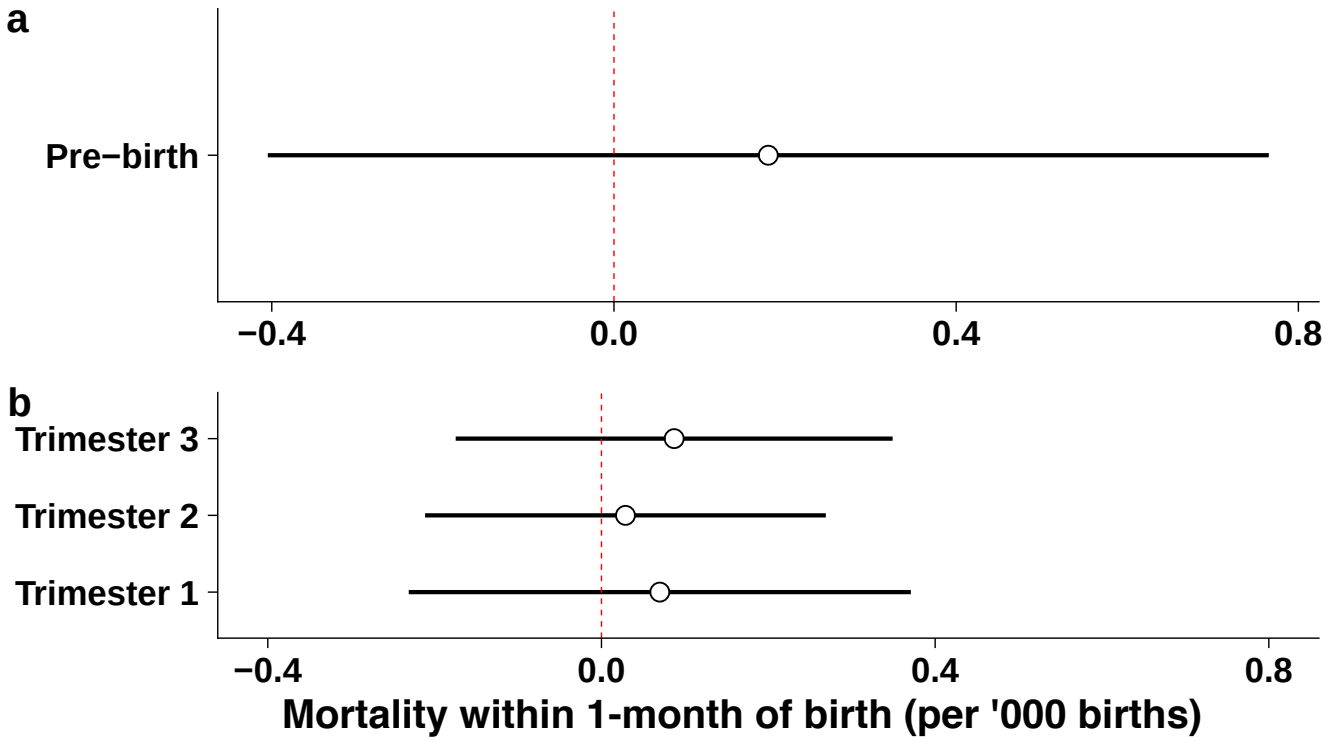

Fig. S4. Regression estimates of pre-birth exposure to outdoor biomass burning on mortality within 1-month of birth. **a** Coefficient estimate on upwind burned area within 0-30 km on mortality within 1-month after birth. **b** Coefficient on trimester-wise upwind burned areas. Circles indicate point estimates, and whiskers the 95% confidence interval on the point estimate. Each plot shows estimates from a separate regression model. Regression specifications are similar to Equation [1] (Methods), but exclude post-birth exposure variables. All other control variables are included in the regressions. **a** shows the estimate from a model using average exposure over the whole pre-birth period. **b** uses a model with average exposure within each trimester of the pre-birth period. Sample used is the DHS births data (N = 2.3 million). The sample mean under 1-month mortality is 36.4 deaths per '000 births.

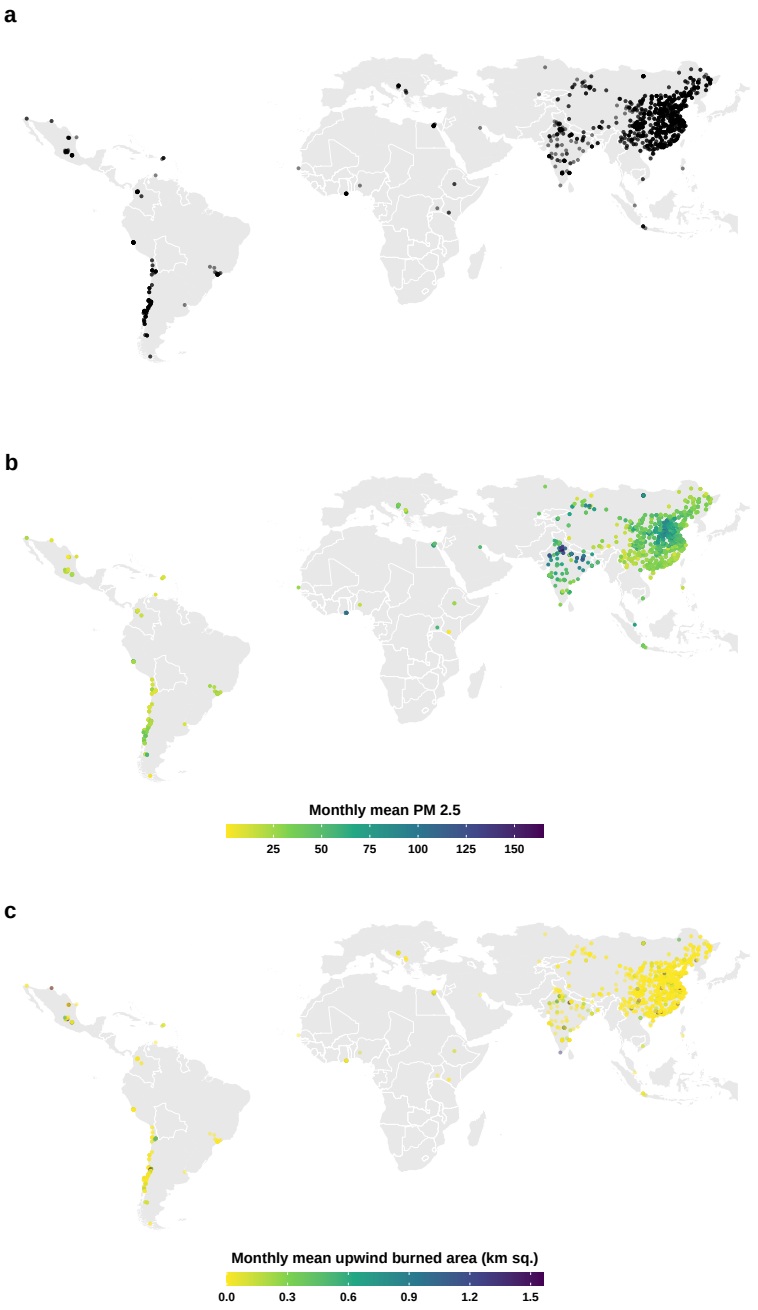

Fig. S5. Location, average $PM_{2.5}$ and burned area exposure at ground station monitors. **a** Location of ground station monitors used for estimation ($N = 2040$). **b** Monthly sample average $PM_{2.5}$ ($\mu g/m^3$) from 2014 - 2018 recorded by monitors. **c** Average upwind exposure to biomass burned area in (square kilometers per month). Exposure is based on monthly burned area recorded in the upwind quadrant within a 30 km distance. The range of values in **b** and **c** are capped at the 99^{th} percentile of the distribution for better visualization. White borders highlight the countries in the DHS births sample used in the infant mortality estimation.

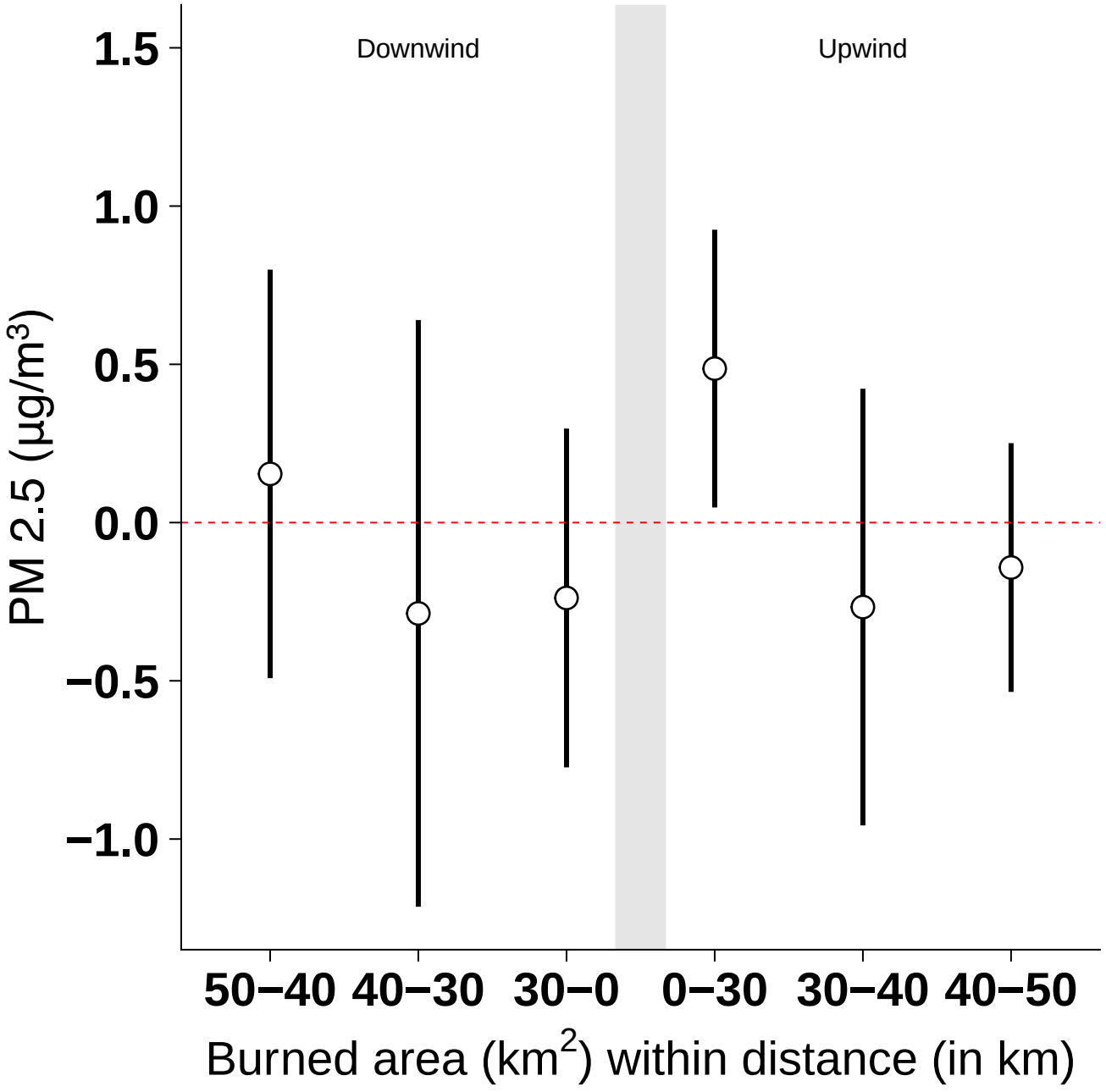

Fig. S6. Regression estimates of outdoor biomass burning on $PM_{2.5}$ at ground station monitors. Plot shows the coefficient estimates from the regression model in Equation [2] (Methods) for the impact of monthly burned area in up and downwind directions around ground station monitors on $PM_{2.5}$ recorded at the monitors. Each additional square kilometer increase in nearby biomass burning (within 30 km) in the upwind direction increases $PM_{2.5}$ by 0.49 $\mu g/m^3$. Biomass burning that is further away (30 to 40 or 40 to 50 km) has no effect. Downwind burned areas do not have a significant effect on $PM_{2.5}$ – consistent with pollution from burning blowing away from the location of the air pollution monitors. Circles indicate point estimates, and whiskers the 95% confidence interval on the point estimate.

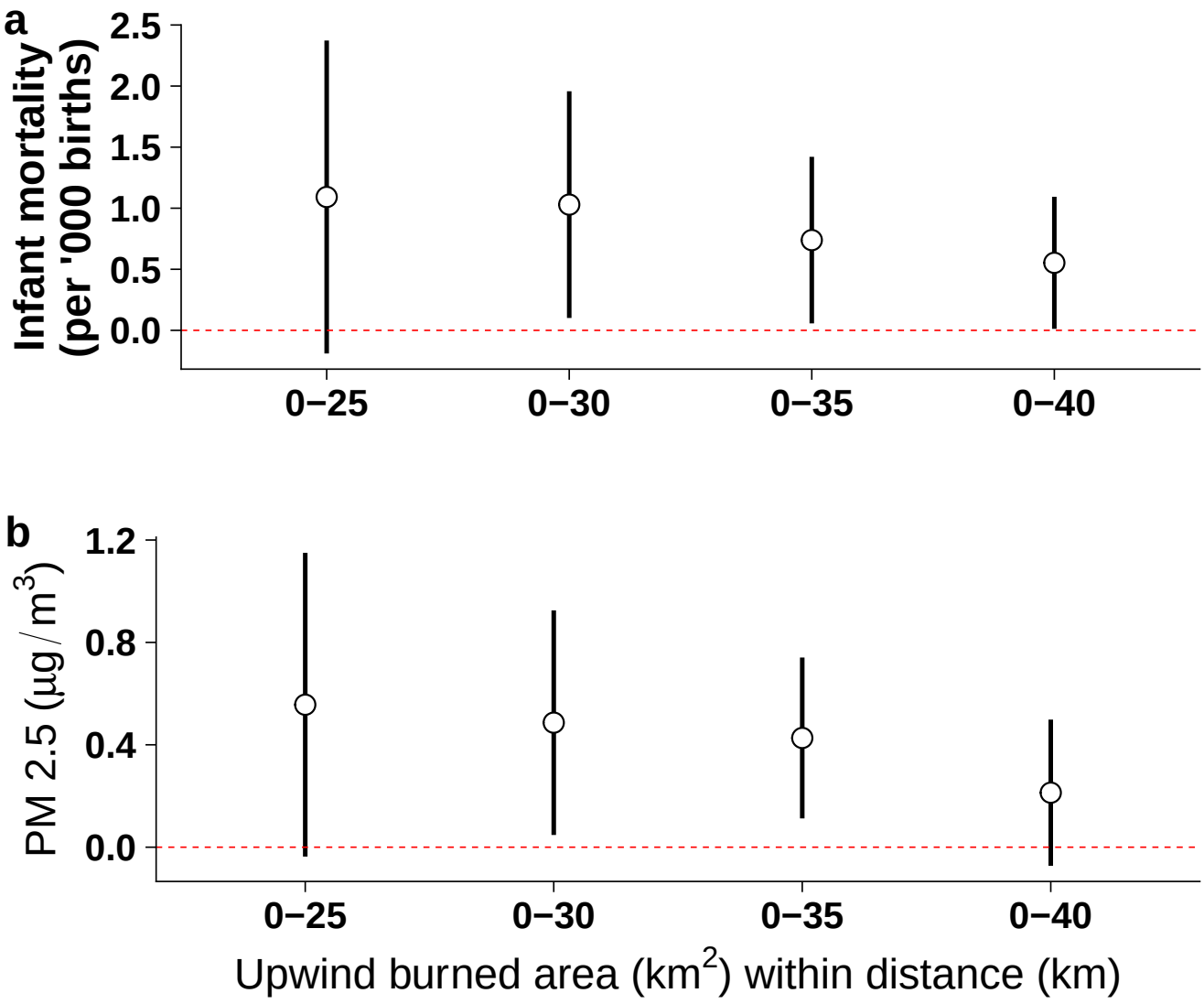

Fig. S7. Variation in the effect of upwind burned area on infant mortality and particulate pollution using different exposure radii. **a** shows the coefficient estimates of the effect of post-birth upwind burned area exposure on infant mortality. Each coefficient is from a separate regression model. Regression specifications are similar to Equation [1], but exclude pre-birth exposure. The effect of each additional square kilometer upwind burned area on infant mortality declines in magnitude as the distance used to define exposure increases. **b** shows the effect of monthly upwind burned area on $PM_{2.5}$ at ground station monitors when the distance used to define nearby exposure is changed. Each estimate is from a separate regression model using specifications similar to Equation [2]. The effect of each additional square kilometer upwind burned area on $PM_{2.5}$ shows a strikingly similar pattern to that seen in **a** as the distance used to define exposure is varied. Circles indicate point estimates, and whiskers the 95% confidence interval on the point estimate.

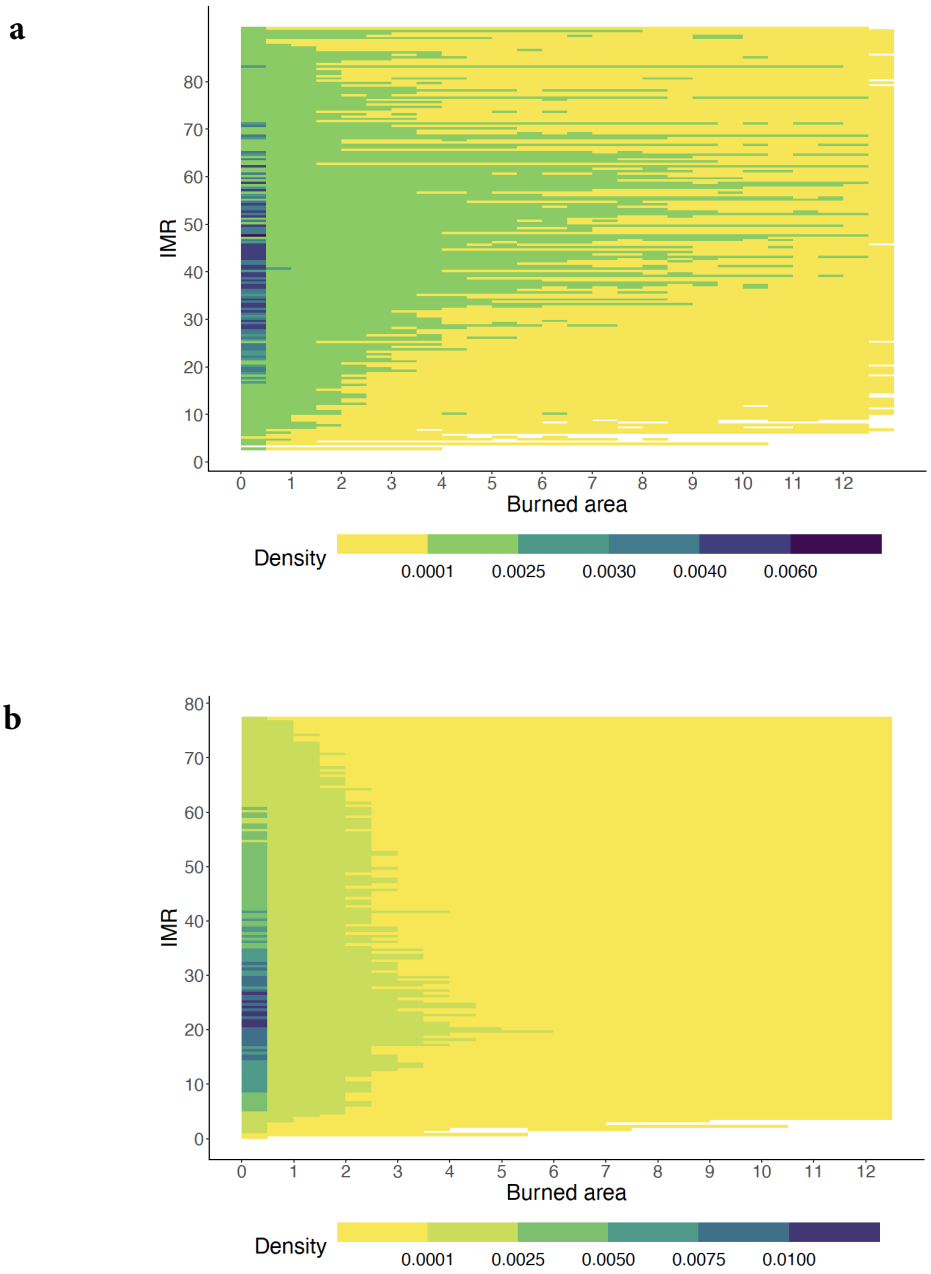

Fig. S8. Distribution of burned area and infant mortality rates in estimation sample and extended sample . Density of infant mortality and burned area distribution in **a** the DHS births data used in the regression estimates, and **b** in the extended sample used for calculating the global number of attributable deaths.

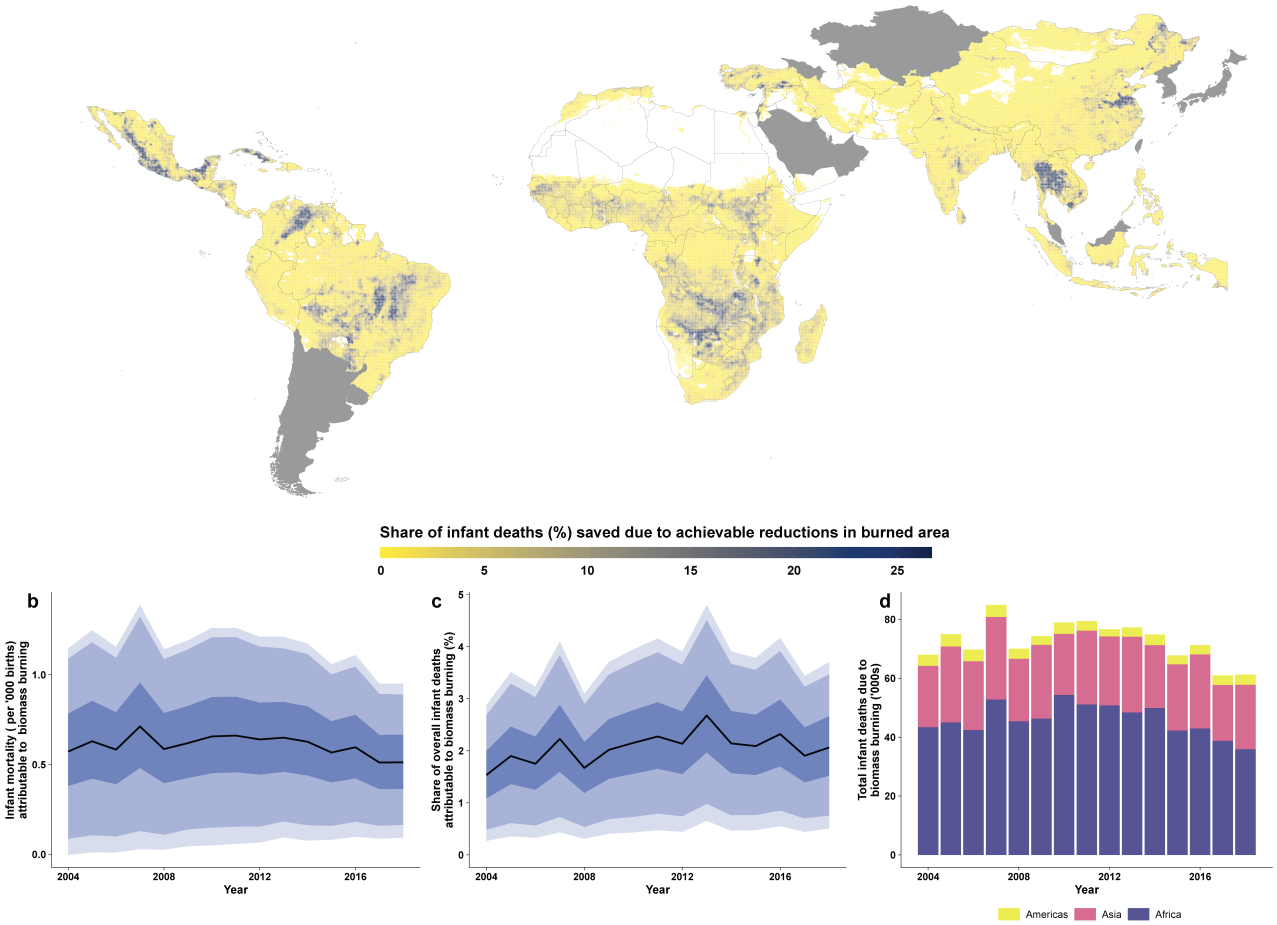

Fig. S9. Avoided infant deaths from achievable reduction in post-birth exposure to outdoor biomass burning. **a** Average share of overall infant deaths avoided if biomass burning was reduced to achievable levels (the minimum observed burned area at each grid cell location during 2004-18) over the period 2004 to 2018. **b**, **c** and **d**,respectively, show the annual trends in **(b)** births-weighted infant mortality (deaths per '000 births) attributable to biomass burning exposure, **(c)** average infant mortality due to biomass burning exposure as share of overall infant mortality (%), and **(d)** number of avoided infant deaths in '000s by region that result from achievable reductions biomass burning. The colors in the stacked bar charts in **d** show the break-up of the total infant deaths across three broad regions in the sample – Africa, Asia and the Americas. The solid lines in **b** and **c** show the sample median, and the shaded regions show the 25^{th} to 75^{th} (darkest), 10^{th} to 90^{th} (medium), and 5^{th} to 95^{th} (lightest) percentile ranges based on bootstrapped estimates of predicted infant mortality values at each 1 km X 1 km grid cell, for each year.

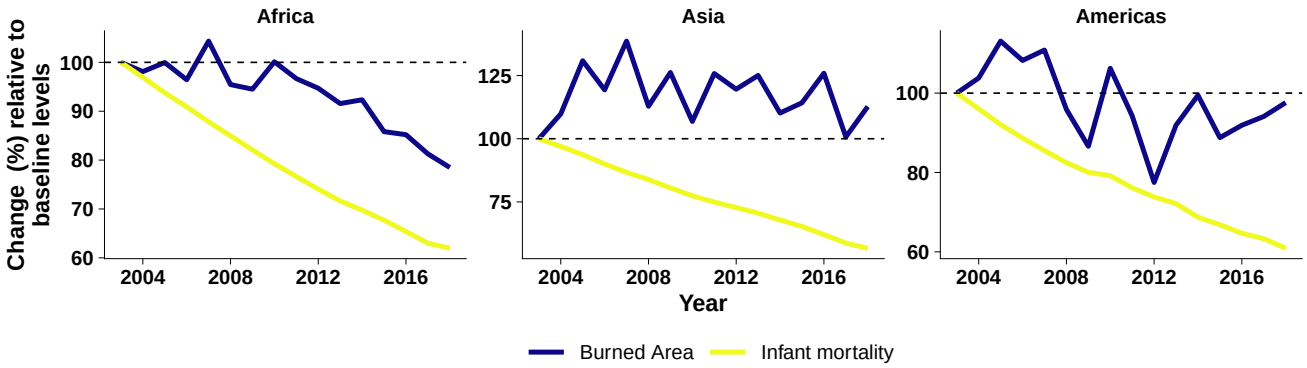

Fig. S10. Births-weighted trends in burned area and infant mortality rate by region, relative to baseline values. Plots show the annual birth-weighted average biomass burned area and infant mortality rate, averaged over the 5 km $\times$ 5 km grid-cells used in the extended global sample. Annual values are normalized setting baseline values to 100. Baseline values are the 3-year average from 2001-2003 for burned area, and 2003 levels for infant mortality.

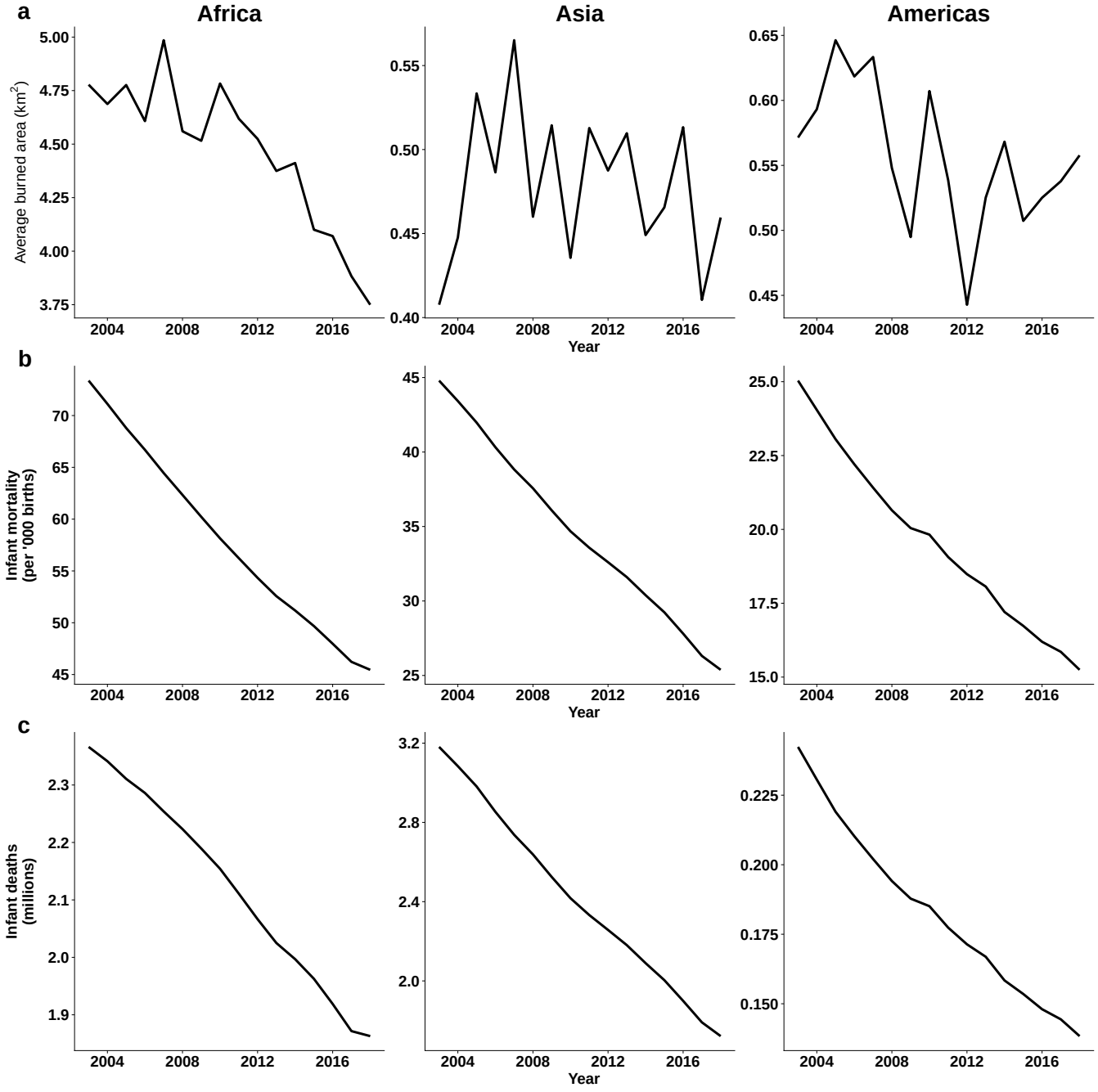

Fig. S11. Trends in births-weighted burned area, births-weighted infant mortality rate, and total infant deaths by region in the extended sample. **a** Annual birth-weighted average upwind biomass burned area and infant mortality rate, averaged over all grid-cells used in the extended sample. **b** Annual infant mortality rate (births-weighted average over all grid cells). **c** Total number of infant deaths estimated as the product of infant mortality rate and number of births, summed up over all grid cells in the region for each year.

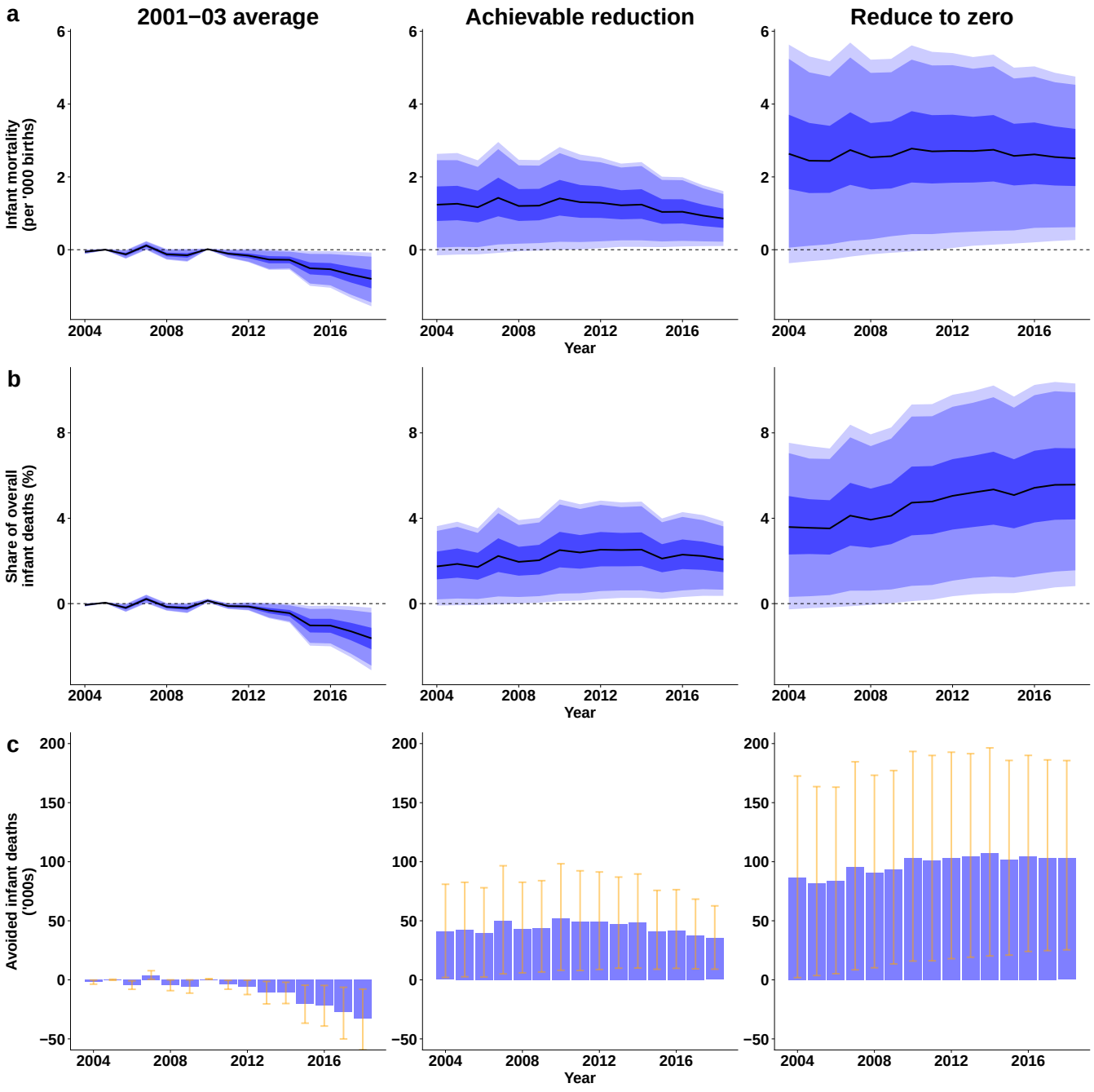

Fig. S12. Avoided infant deaths in Africa region from reduced post-birth exposure to outdoor biomass burning under different scenarios. **a**, **b**, and **c**, respectively, show infant mortality attributable to biomass burning exposure, share of overall infant mortality (%), and number of avoided infant deaths for three scenarios, from left to right – burning held at the baseline values, reduced to achievable levels (the minimum observed burned area at each grid cell location during 2004-18), and complete reduction. Shaded regions in **a** and **b** show the 25^{th} to 75^{th} (darkest), 10^{th} to 90^{th} (medium), and 5^{th} to 95^{th} (lightest) percentile ranges based on bootstrapped estimates of predicted infant mortality values at each 1 km X 1 km grid cell, for each year. Error bars in **c** show 5^{th} to 95^{th} percentile range and the bar height represents the median.

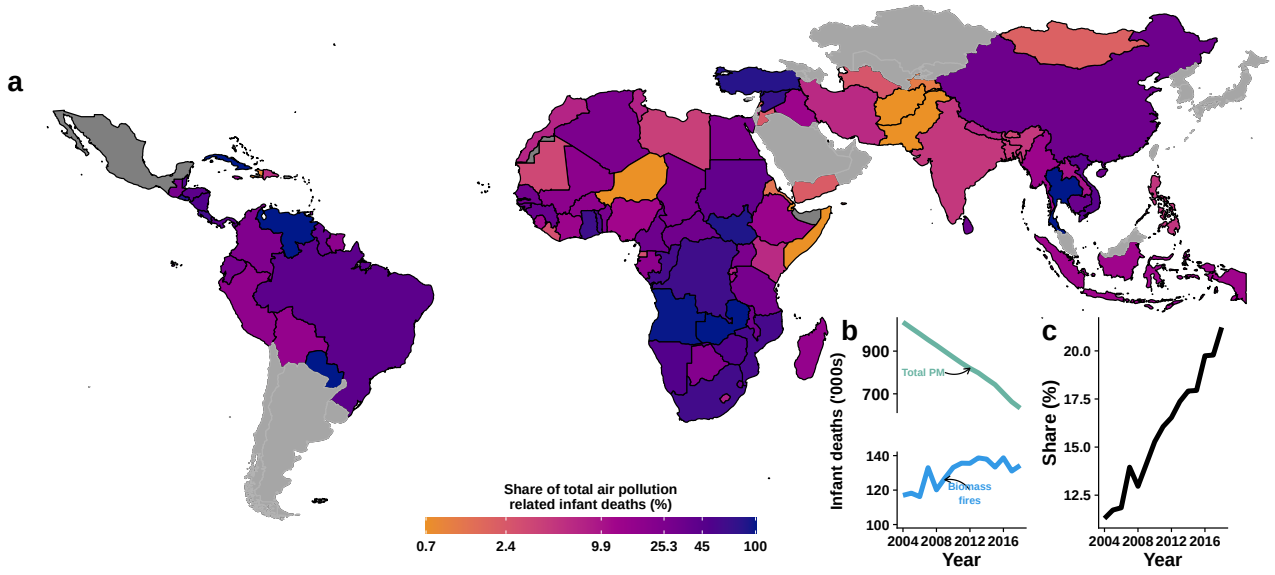

Fig. S13. Share of total air pollution-related infant deaths attributable to biomass fires exposure. **a** Share of total infant deaths due to particulate matter pollution (PM) attributable to biomass burning exposure estimated in this study (averaged over the study period 2004 - 2018 within each country in sample). **b** Yearly trend in annual infant deaths attributable to overall PM and the estimated infant deaths due to exposure to biomass fires (annual total across all sample countries). **c** Annual trend in the share of biomass fire exposure in overall PM infant deaths (average across all countries for each year weighted by total PM infant deaths). Overall PM-related infant deaths are based on Global Burden of Disease (GBD) estimates (1), calculated as deaths occurring in early, late, or post neonatal age groups due to particulate matter pollution risk. Infant deaths due to biomass fires are estimated by aggregating the grid-cell level estimates from this study to the country-year level.

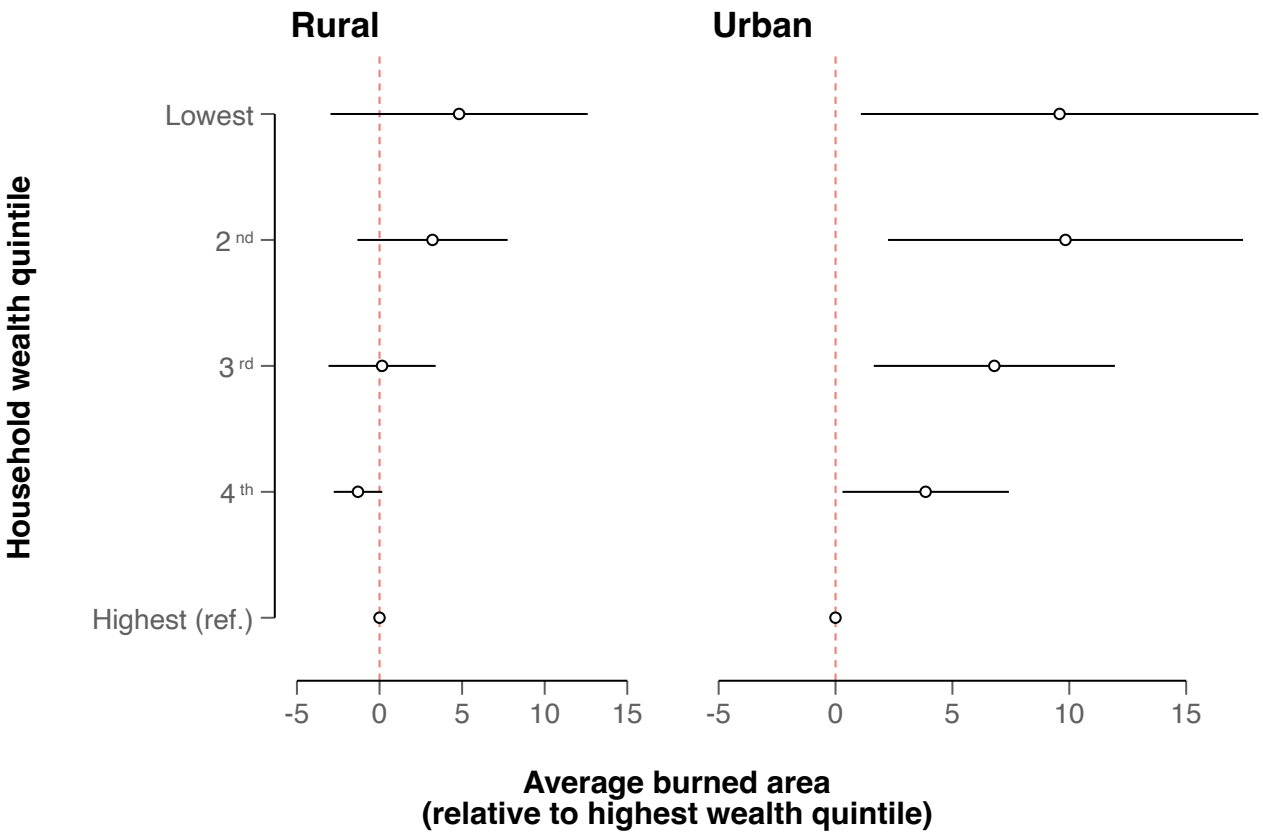

Fig. S14. Cross-sectional relationship between total burned area and household wealth level in rural and urban DHS clusters. We regress the average monthly biomass burned area in the two years preceding the survey year (within a 30-km radius around the households' cluster location) on household wealth quintile indicators, controlling for country and survey year fixed effects.

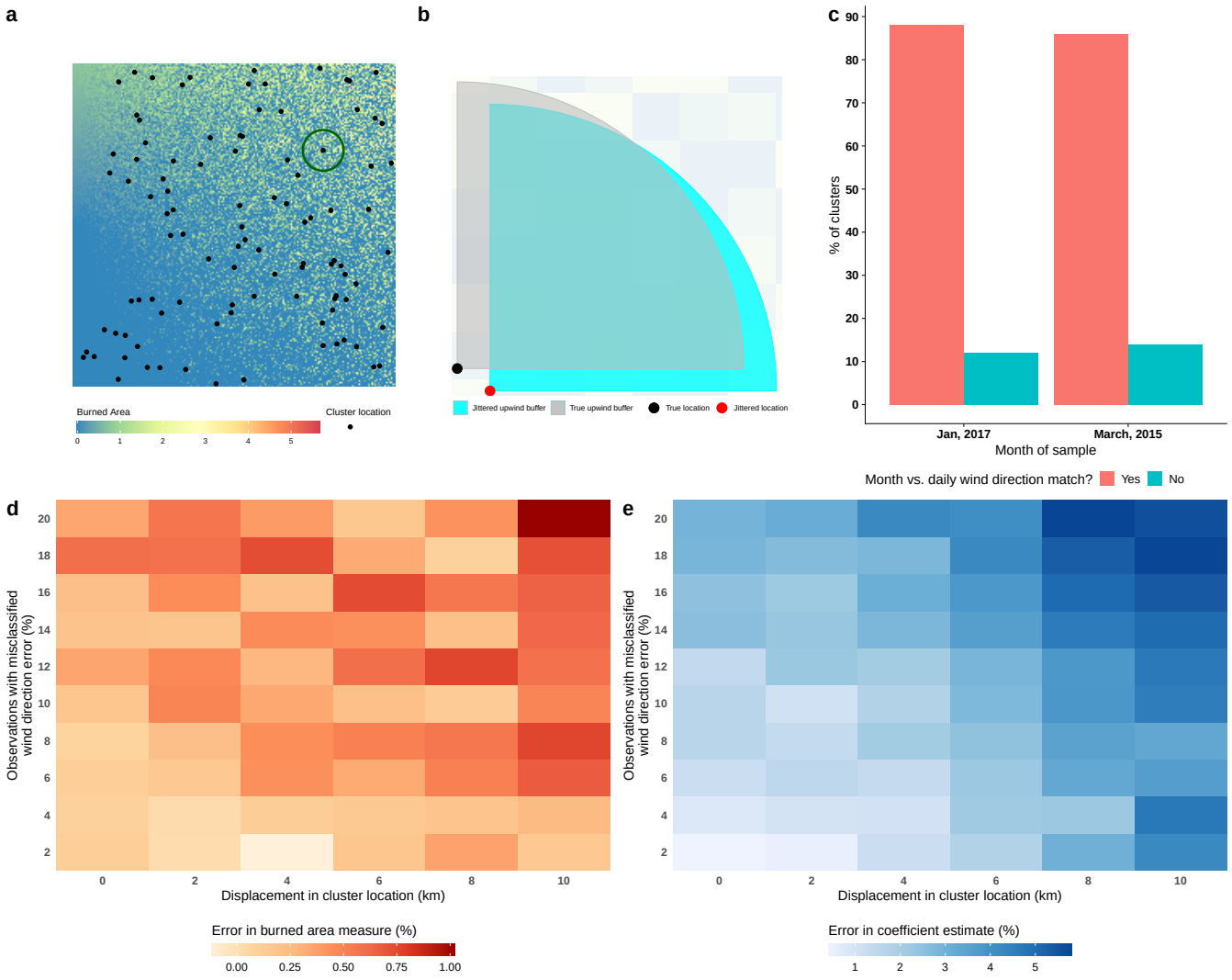

Fig. S15. Simulating the role of measurement error in DHS cluster location and wind direction. **a** shows the simulated burned area with 100 randomly drawn locations (dark dots) representing DHS clusters. **b** zooms in on a particular cluster marked by the red circle in **a** and shows the random displacement in the location and the associated shift in the buffer used to measure upwind burned area. **c** shows the degree of mismatch in upwind direction when using monthly average wind direction versus the mode of daily wind direction for the month. The comparison is shown for two different random samples of 100 DHS clusters for two different months in the sample period. **d** and **e** show the consequences of measurement error in the upwind burned area and the regression coefficient arising due to displacement of DHS clusters and misclassification of upwind quadrant using a simulation exercise. Simulation is carried out by drawing a sample of 100 points with 100 repetitions for each displacement – wind-direction error combination (Methods). At the extreme, if DHS clusters were randomly displaced by 10 km and 20% of them had wrong upwind quadrants, upwind burned area would contain a 1% error, and the regression coefficient would be biased up to 5% from the true effect.

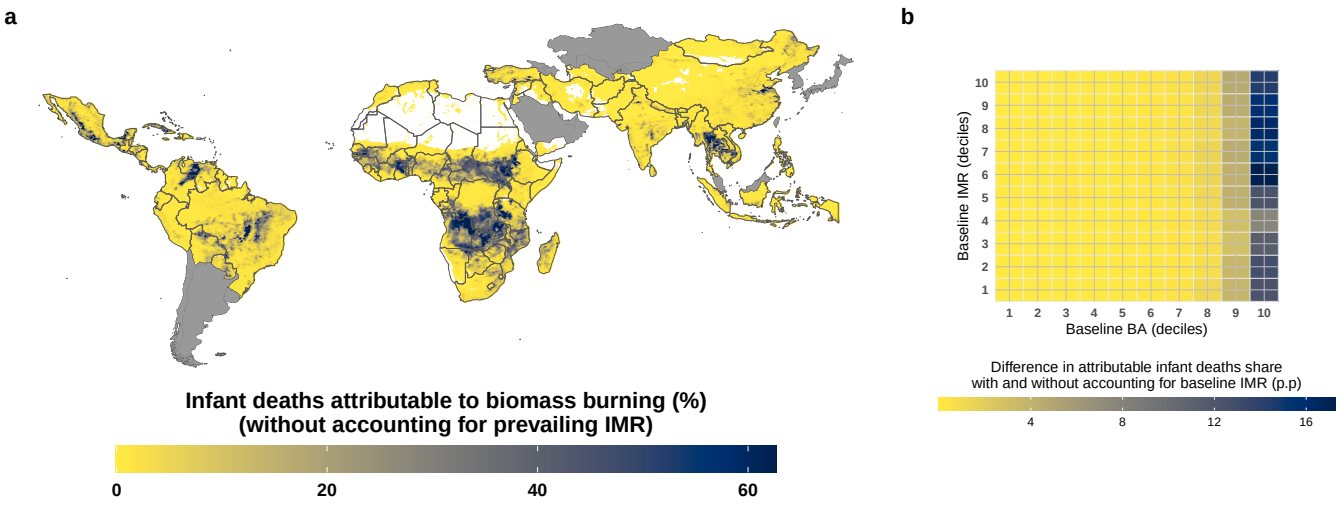

Fig. S16. Avoided infant deaths from eliminating outdoor biomass burning without incorporating prevailing baseline IMR. **a** Average share of overall infant deaths avoided if biomass burning was reduced to zero over the period 2004 to 2018 estimated using a model that excludes the moderating effect of prevailing baseline infant mortality rates. **b** The difference in the share of overall infant deaths avoided estimated from the model with no prevailing IMR modifier versus that accounting for the moderating effect of prevailing IMR (shown in Fig 4a). The difference is expressed in percentage points (p.p). **b** shows this difference as a heatmap with deciles of baseline (2003) burned area on the horizontal axis and deciles of baseline IMR on the vertical axis. Overall, the difference is 2.16 percentage points on average (median of 0.16).

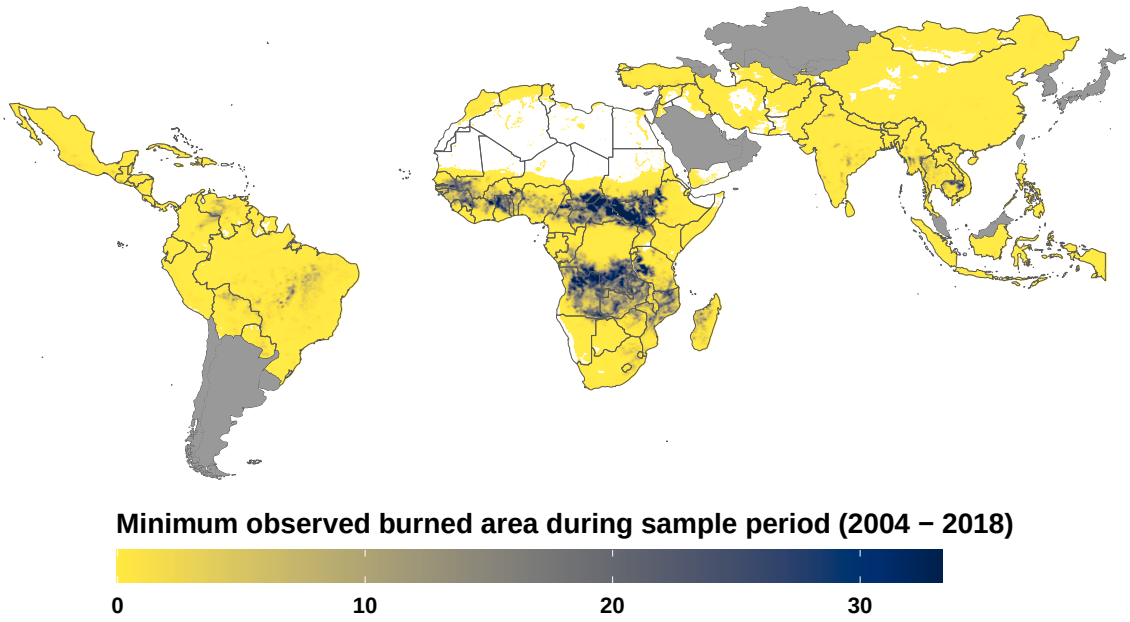

Fig. S17. Plausibly achievable reduction in burned area. Map shows the minimum observed burned area during the period 2004 to 2018 at each location (in kilometer square). These minimum observed values are used as the counterfactual in the "achievable reduction" in burned area scenario.

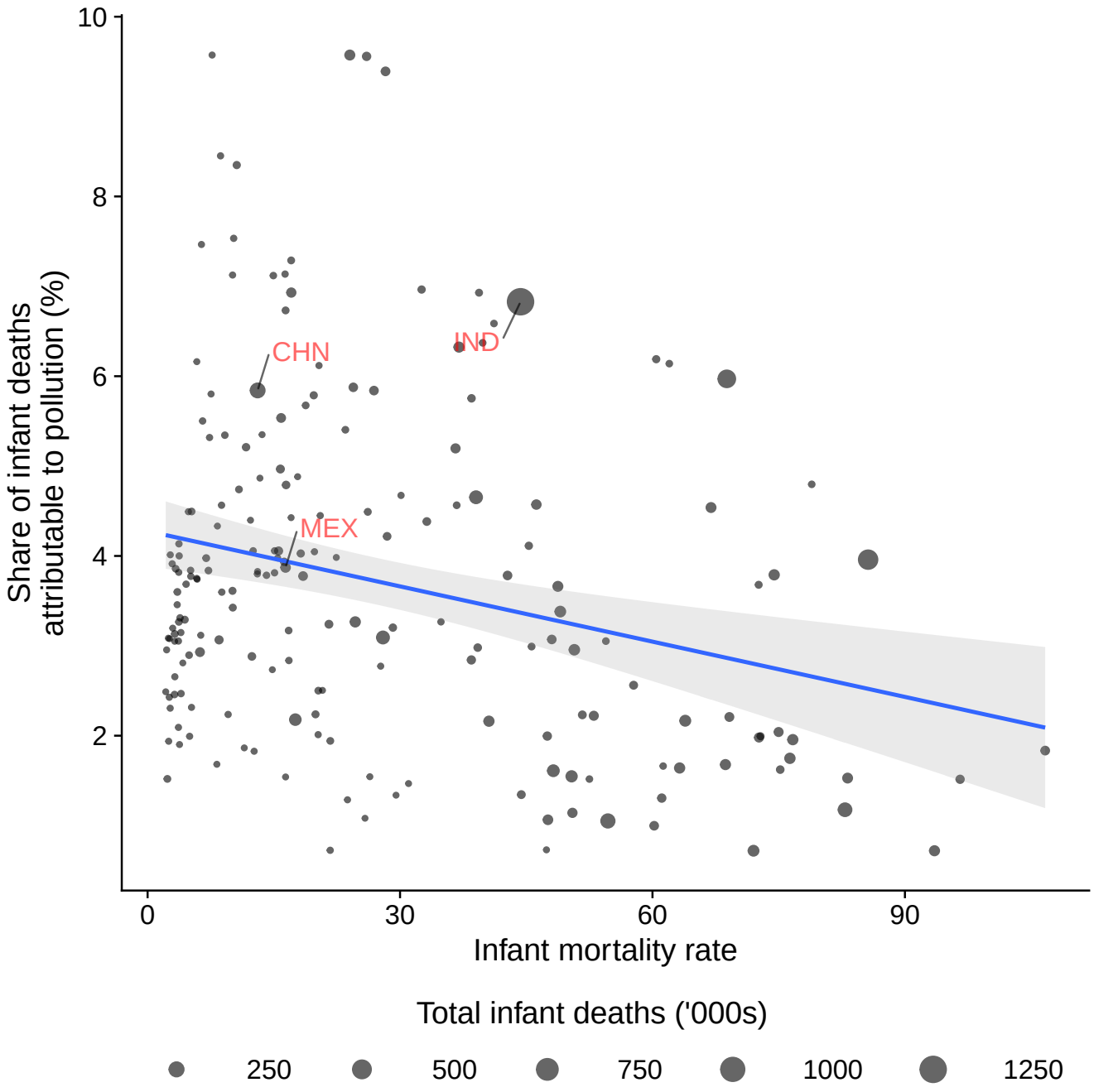

Fig. S18. Cross-country relationship between share of infant deaths due to pollution and infant mortality rate. Plot shows a scatter plot of the average (2003 - 2018) infant mortality rate on the horizontal axis and the share of infant deaths attributable to ambient $PM_{2.5}$ pollution on the vertical axis, along with the linear fit line for a sample of 193 countries. The size of the bubbles is proportional to the total number of infant deaths per year. Infant mortality rates are drawn from the World Development Indicators database (2), while the total number of infant deaths and those attributable to ambient air pollution are based on the Global Burden of Disease estimates (1).

Table S1. List of DHS surveys in sample

Country	Survey Year(s)	Observations
Albania	2008, 2017	10019
Angola	2006, 2011, 2015	41300
Armenia	2010, 2015	6198
Bangladesh	2004, 2007, 2011, 2014, 2018	61822
Benin	2012, 2017	53518
Bolivia	2008	7203
Burkina Faso	2010	19339
Burundi	2010, 2016	41820
Cambodia	2005, 2010, 2014	29942
Cameroon	2004, 2011, 2018	40862
Chad	2014	41791
Colombia	2010	22678
Comoros	2012	5197
Cote d'Ivoire	2012	12052
Democratic Republic of the Congo	2007, 2013	37141
Dominican Republic	2007, 2013	14218
Egypt	2005, 2008, 2014	43418
Ethiopia	2005, 2010, 2016	44121
Gabon	2012	9358
Ghana	2008, 2014	14971
Guinea	2005, 2012, 2018	34729
Guyana	2009	2145
Haiti	2006, 2012, 2016	30927
Honduras	2011	16612
India	2015	636579
Jordan	2007, 2012, 2017	55610
Kenya	2008, 2014	50950
Kyrgyz Republic	2012	6732
Lesotho	2004, 2009, 2014	11706
Liberia	2007, 2009, 2013	22349
Madagascar	2008	12162
Malawi	2004, 2010, 2015	69101
Mali	2006, 2012, 2018	47963
Moldova	2005	519
Morocco	2003	14
Mozambique	2011	16142
Myanmar	2015	12121
Namibia	2006, 2013	12087
Nepal	2006, 2011, 2016	24257
Nigeria	2008, 2010, 2013, 2018	175957
Pakistan	2006, 2017	33279
Philippines	2008, 2017	34351
Rwanda	2005, 2008, 2010, 2014	37129
Senegal	2005, 2008, 2010	35582
Sierra Leone	2008, 2013, 2019	54215
South Africa	2016	8159
Swaziland	2006	1651
Tajikistan	2012, 2017	21860
Tanzania	2007, 2010, 2015	37567
Timor	2009, 2016	29852
Togo	2013	14010
Uganda	2006, 2009, 2011, 2016	56676
Zambia	2007, 2013, 2018	55338
Zimbabwe	2005, 2010, 2015	22006

Table S2. Regression results for main specification and robustness of post-birth exposure to alternative specifications

Dependent Variable:	Infant mortality (per '000 births)					
	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Upwind exposure, post-birth	1.0601**	1.1841**	0.9553**	1.0382*	1.0672**	1.2147**
	(0.5322)	(0.5026)	(0.4860)	(0.5408)	(0.5333)	(0.5951)
Downwind exposure, post-birth	0.0018	0.2519	0.1144	-0.1486	0.0101	0.0244
	(0.4932)	(0.4735)	(0.4535)	(0.4925)	(0.4953)	(0.5594)
Upwind exposure, pre-birth	-0.1531	-0.3802	-0.1374	-0.1314	-0.1693	-0.1071
	(0.4102)	(0.3483)	(0.3580)	(0.4162)	(0.4105)	(0.4895)
Downwind exposure, pre-birth	0.1581	0.2191	0.0578	0.1905	0.1511	0.2058
	(0.4156)	(0.3619)	(0.4053)	(0.4213)	(0.4165)	(0.5059)
Non up/downwind exposure, post-birth						-0.1077
						(0.5000)
Non up/downwind exposure, pre-birth						-0.4598
						(0.3719)
Postbirth Upwind - Downwind						
Estimate	1.0583	0.9322	0.8410	1.1868	1.0571	1.1903
std. error	0.6438	0.5637	0.5592	0.6503	0.6451	0.6530
p-value	0.1002	0.0982	0.1326	0.0680	0.1013	0.0683
Fixed-effects						
DHS sample cluster	Yes	Yes	Yes	Yes	Yes	Yes
Country-Birth month	Yes			Yes	Yes	Yes
Country-Birth year	Yes			Yes	Yes	Yes
1-degree grid cell-Birth month		Yes				
1-degree grid cell-Birth year		Yes				
2-degree grid cell-Birth month			Yes			
2-degree grid cell-Birth year			Yes			
Observations	2,237,305	2,237,305	2,237,305	2,237,305	2,237,305	2,237,305

Table shows estimates from separate regressions in each column. Regression models are based on the specification shown in Equation [1]. Up and downwind exposure refer to the outdoor biomass burned area within 0-30 km in the up and downwind quadrants (Fig S2), measured as the average monthly burned area in square kilometers during the nine months preceding birth (pre-birth) and the 12 months after birth, including month of birth (post-birth). Column **1** shows the results from the main specification with country by birth month and country by birth year fixed effects. Columns **2** and **3** flexibly control for seasonal and year effects at a finer spatial scale using 1-degree and 2-degree grid cell fixed effects, respectively, instead of country-level fixed effects. Column **4** shows the estimates without child, maternal, and household control variables. Column **5** excludes climatic factors (temperature, precipitation, and wind speed) from the model. Column **6** adds controls for burned area that is not in either the upwind or downwind directions. For all columns, the estimated difference between the coefficients on postbirth upwind and downwind exposure is shown under *Postbirth Upwind - Downwind* along with the associated standard errors and p-values Values in parentheses show the standard errors clustered at 1-degree grid cell level. Coefficient significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table S3. Regression results limiting sample to births that have occurred since the respondent has resided in the current survey location

Dependent Variable:	Infant mortality (per '000 births)	
	(1)	(2)
Variables		
Upwind exposure, post-birth	1.375**	1.183*
	(0.6306)	(0.7083)
Downwind exposure, post-birth	-0.0039	-0.2848
	(0.6325)	(0.6893)
Upwind exposure, pre-birth	-0.0756	-0.2547
	(0.4694)	(0.4926)
Downwind exposure, pre-birth	0.7967**	0.7470*
	(0.4025)	(0.4419)
Fixed-effects		
DHS sample cluster	Yes	Yes
Country-Birth month	Yes	Yes
Country-Birth year	Yes	Yes
Observations	1,506,527	1,336,201

Table shows estimates from separate regressions in each column. Regression models are based on the specification shown in Equation [1]. Up and downwind exposure refer to the outdoor biomass burned area within 0-30 km in the up and downwind quadrants (Fig S2), measured as the average monthly burned area in square kilometers during the nine months preceding birth (pre-birth) and the 12 months after birth, including month of birth (post-birth). We limit the sample to births that have occurred to mothers' residing in the survey location since before the birth occurred. To do so, we use information from the variable "how long has the respondent lived in the current place of residence" and exclude births that occurred before the respondent moved to the current location. This information is unavailable for 48 of the 118 DHS surveys used in our analysis. From the remaining 1,506,527 births, 1,336,201 births occurred to mothers who had resided in the current location before the birth occurred. Column **1** shows the estimates using our main specification for the sub-sample of births for which the location history is available. In column **2**, we present estimates for the 1,336,201 births that occurred to mothers who had resided in the current location before the birth occurred. Values in parentheses show the standard errors clustered at 1-degree grid cell level. Coefficient significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

Table S4. Regression results - heterogeneity in the effect of post-birth upwind burned area by household wealth, baseline infant mortality, and baseline particulate pollution

Dependent Variable:	Infant mortality (per '000 births)		
	(1)	(2)	(3)
Variables			
Upwind BA × Wealth Quintile 1	0.8850		
	(0.5913)		
Upwind BA × Wealth Quintile 2	1.274**		
	(0.6070)		
Upwind BA × Wealth Quintile 3	1.203**		
	(0.5813)		
Upwind BA × Wealth Quintile 4	0.9109		
	(0.8028)		
Upwind BA × Wealth Quintile 5	0.7809		
	(0.7829)		
Upwind BA × IMR		-0.0054***	
		(0.0018)	
Upwind BA × $PM_{2.5}$			-0.0046
			(0.0129)
Fixed-effects			
DHS sample cluster	Yes	Yes	Yes
Country-Birth month	Yes	Yes	Yes
Country-Birth year	Yes	Yes	Yes
Wealth quintile	Yes		
Observations	2,237,307	2,237,307	2,237,307

Table shows estimates from separate regressions in each column. Regression models are based on the specification shown in Equation **??**. "Upwind BA" refers to average monthly upwind outdoor biomass burned area exposure (in km^2) within 0-30 km distance during the post-birth period. For brevity, each column shows the coefficients on the interaction of post-birth upwind burned area exposure with the modifiers household wealth in column **1**, baseline infant mortality rate in column **2**, and baseline pollution ($PM_{2.5}$) in column **3**. Baseline IMR is constructed as follows: we take the sample infant mortality rate for the year prior to birth averaged over clusters located within 1-degree grid-cells around each birth location. Baseline $PM_{2.5}$ is similarly constructed as the lagged average $PM_{2.5}$ at 1-degree grid-cells around each birth location. Values in parentheses show the standard errors clustered at the DHS sample cluster level. Coefficient significance at 1%, 5% and 10% are indicated by ***, ** and *, respectively.

References

1. GB of Disease Collaborative Network, Global burden of disease study 2017 (gbd 2017) results. *Seattle, United States: Inst. for Heal. Metrics Eval. (IHME)* (2018).
2. W Bank, *World development indicators*. (The World Bank), (2022).