Mortality Associated with Ambient PM$_{2.5}$ Exposure in India: Results from the Million Death Study

Patrick E. Brown,† Yurie Izawa,† Kalpana Balakrishnan,‡ Sze Hang Fu,† Joy Chakma,§ Geetha Menon,∥ Rajesh Dikshit,‡ R.S. Dhaliwal,‡ Peter S. Rodriguez,§ Guowen Huang,∥ Rehana Begum,§ Howard Hu,¶ George D’Souza,** Randeep Guleria,* and Prabhjot Jha***

1Centre for Global Health Research (CGHR), St Michael’s Hospital and Dalla Lana School of Public Health, The University of Toronto, Toronto, Ontario, Canada
2Department of Environmental Health Engineering, Sri Ramachandra Institute of Higher Education and Research, Chennai, India
3The Indian Council of Medical Research, New Delhi, India
4Centre for Cancer Epidemiology, Tata Memorial Centre, Mumbai, India
5Department of Preventive Medicine, Keck School of Medicine of University of Southern California, Los Angeles, USA
6St. John’s Medical College, St. John’s Research Institute, Bangalore, India
7All India Institutes of Medical Sciences, New Delhi, India

BACKGROUND: Studies on the extent to which long-term exposure to ambient particulate matter (PM) with aerodynamic diameter $\leq 2.5$ μm (PM$_{2.5}$) contributes to adult mortality in India are few, despite over 99% of Indians being exposed to levels that the World Health Organization (WHO) considers unsafe.

OBJECTIVE: We conducted a retrospective cohort study within the Million Death Study (MDS) to provide the first-ever quantification of national mortality from exposure to PM$_{2.5}$ in India from 1999 to 2014.

METHODS: We calculated relative risks (RRs) by linking a total of ten 3-y intervals of satellite-based estimated PM$_{2.5}$ exposure to deaths 3 to 5 y later in over 7,400 small villages or urban blocks covering a total population of 6.8 million. We applied using a model-based geostatistical model, adjusted for individual age, sex, and year of death; smoking prevalence, rural/urban residency, area-level female illiteracy, languages, and spatial clustering and unit-level variation.

RESULTS: PM$_{2.5}$ exposure levels increased from 1999 to 2014, particularly in central and eastern India. Among 212,573 deaths at ages 15–69 y, after spatial adjustment, we found a significant RR of 1.09 [95% credible interval (CI): 1.04, 1.14] for stroke deaths per 10-μg/m$^3$ increase in PM$_{2.5}$ exposure, but no significant excess for deaths from chronic respiratory disease and ischemic heart disease (IHD), all nonaccidental causes, and total mortality (after excluding stroke). Spatial adjustment attenuated the RRs for chronic respiratory disease and IHD but raised those for stroke. The RRs were consistent in various sensitivity analyses with spatial adjustment, including stratifying by levels of solid fuel exposure, by sex, and by age group, adjustment of climatic variables, and in supplementary case–control analyses using injury deaths as controls.

DISCUSSION: Direct epidemiological measurements, despite inherent limitations, yielded associations between mortality and long-term PM$_{2.5}$ exposure. Our findings are consistent with those reported in earlier models used by the WHO to derive estimates of PM$_{2.5}$ mortality in India. The modest RRs in our study are consistent with near or null mortality effects. They suggest suitable caution in estimating deaths from PM$_{2.5}$ exposure based on MDS results and even more caution in extrapolating model-based associations of risk derived mostly from high-income countries to India. https://doi.org/10.1289/EHP9538

Introduction

Epidemiological studies of the health consequences of exposure to particulate matter (PM) with aerodynamic diameter $\leq 2.5$ μm (PM$_{2.5}$) have been conducted mostly in high-income countries with lower levels of exposure than those in India. In 2014, average Indian exposure to PM$_{2.5}$ was 47 μg/m$^3$ based on satellite-derived estimates (Brauer et al. 2016), with over 99% of India’s population exposed to levels above the ceiling of 10 μg/m$^3$ recommended as “safe” by the World Health Organization (WHO) (WHO 2006). In India, household solid cooking fuels contribute to 20%–50% of ambient PM$_{2.5}$ exposure, a much higher proportion than in other countries (Chowdhury et al. 2019). With such high levels of exposure, even modest excess mortality caused by PM$_{2.5}$ would result in a substantial number of deaths. Indian hazards likely differ substantially from those studied in high-income countries given distinct disease patterns, including high age-standardized death rates for ischemic heart disease (IHD) and stroke (Ke et al. 2018).

In the absence of nationally representative evidence on the extent to which PM$_{2.5}$ exposure increases cause-specific mortality rates in India, the WHO and others have relied on modeled associations that apply various complex transformations to hazards observed in studies mainly in high-income countries (Burnett et al. 2018). The WHO thus estimates that 1.1 million Indians of all ages died from exposure to ambient air pollution in 2016, to which PM$_{2.5}$ is the major contributor (WHO 2018). In these model-based death totals, about three-quarters were from respiratory and vascular disease in adults. Here, we conducted a retrospective cohort study within the Million Death Study (MDS) in India to quantify the association of cause-specific adult mortality with long-term exposure to PM$_{2.5}$. We focused on quantification of the risks specific to the Indian population and considered misclassification of exposures and diseases. Given that many previous assessments have not appropriately considered clustering of exposure and disease, we pay specific attention to showing the variation in estimated risks with and without spatial adjustment.

Methods

Satellite-Based PM$_{2.5}$ Estimates

We obtained satellite-derived modeled estimates of ambient PM$_{2.5}$ data from 1998 to 2014 from van Donkelaar et al. (2015). Briefly,
as described by Boys et al. (2014), van Donkelaar et al. (2015) combined data from three types of satellite instruments together with the GEOS-Chem chemical transport model (version 9-01-03; http://geos-chem.org) to estimate global surface PM$_{2.5}$ concentrations at a resolution of 0.1 degrees (approximately 11 km at the equator) for the period 1998–2012. The satellite-based data do not distinguish between sources of PM$_{2.5}$, which differ substantially across countries (e.g., emissions from motor vehicles, industrial facilities, household solid fuel use) and within India [e.g., in northern India during seasons of crop burning, and in rural areas with high levels of use of household solid fuels; Balakrishnan et al. (2013)]. van Donkelaar et al. (2015) derived 3-y medians of the annual satellite-based PM$_{2.5}$ grids to reduce noise in the annual satellite-derived values. In addition, van Donkelaar et al. (2015) evaluated their 3-y medians of PM$_{2.5}$ satellite-derived estimates using data from established ground-level monitoring networks in Canada, the United States, and Europe and collected 210 ground-based observations from the literature for areas outside these regions. Correlations between satellite-based estimates and ground-level data were stronger with more prolonged measurements. For example, for ground data collected for a year or less, the correlation coefficient, $R$, was 0.67, whereas $R$ was 0.79 for ground-level data collected over 10 y. Correlations varied also by region: for Canada and the United States, the $R$ was 0.76; for Europe, the $R$ was 0.73; and for the rest of the world the $R$ was 0.81.

With respect to India, van Donkelaar et al. (2015) evaluated their PM$_{2.5}$ estimates using 55 different locations (representing >250 points in time) distributed over 24 major cities (Figure S1). Although some high population density areas were not represented, all Indian mega-cities (population >4 million) except Surat were represented. Using New Delhi as an example, the authors reported a local underestimation in annual mean PM$_{2.5}$ in comparison with ground observations that seemed to be amplified during the winter season (Figure S2C). Similarly, the authors found that the mean (population-weighted) PM$_{2.5}$ concentration for South Asia was 34.6 µg/m$^3$ (standard deviation (SD) = 15.8) for the period 2001–2010 and that concentrations increased by 2.9%/year [95% confidence interval (CI): 2.2%, 3.6%] for this region in the period 1998–2012.

We further compared alternative satellite PM$_{2.5}$ data from the public domain to ground monitoring data sources from various locations and times (Supplemental Material, “Comparison of PM$_{2.5}$ Data Sources”). Table S1 lists these satellite-derived PM$_{2.5}$ data sources together with a few ground-level sources, and Figure S3 displays the various satellite-derived PM$_{2.5}$ data for the year 2010. We found that the data set by van Donkelaar et al. (2015) has the highest correlation (0.88; $p < 0.001$) with January ground-level data (which is the peak month for PM$_{2.5}$ exposure) and the second highest correlation with the ground-level annual averages (Figure S2A,B).

The MDS

Full details of the methods, strengths, and limitations of the MDS have been published earlier (Aleksandrowicz et al. 2014; Dikshit et al. 2012; Menon et al. 2019; Million Death Study Collaborators 2017). Briefly, in collaboration with the Registrar General of India, the MDS monitored approximately 14 million people in 2.4 million nationally representative households in India from 1998 through 2013 in the Sample Registration System (SRS) (RGI and Census Commissioner 2011a). The sample frames comprised 6,671 (2001–2003) and 7,597 (2004–2013) sampling units, typically villages or small census urban blocks, drawn from the preceding 1991 and 2001 censuses, respectively. Nationally representative demographic statistics can be derived from each sampling frame. Each sampling unit undergoes population enumeration at the beginning of each 10-y period. We used the 2004–2013 SRS sample frame to assign PM$_{2.5}$ estimates and deaths because we had population denominators of the sampling units. Each sampling unit comprised ~150–300 households. The SRS follows the population within these sampling units prospectively for a 10-y period for any deaths. Every 6 months, ~900 full-time, trained, non-medical surveyors update the population composition in these households, capture the date of death, and investigate each death using a well-validated verbal autopsy instrument [based on the WHO (2012) instrument, including a detailed two-page structured list of symptoms and a half-page local language narrative of the circumstances, symptoms, and treatments, if any, prior to death] (Aleksandrowicz 2014). Each field report of a death is converted to an electronic form and randomly assigned to 2 of 404 specially trained physicians. The two physicians independently assigned a cause of death based on the nonmedical surveyors’ field report of the death, using the WHO’s International Classification of Diseases-10th revision (ICD-10) (WHO 2004), according to the cause-of-death assignment manual specifically designed for the MDS (RGI, SRS Collaborators, and CGHR 2011; Menon et al. 2019). Where the two diagnoses did not agree, they underwent anonymous reconciliation by the same two physicians, and a third senior physician anonymously adjudicated persisting differences (which occurred ~18% of the time; Table S2). The percentage of deaths requiring adjudication differed little by sex, urban-rural setting, geographic region, or PM$_{2.5}$ exposure level but was higher among people older than 69 y of age (Table S2). Various quality control procedures resulted in generally low levels of ill-defined deaths before age 70 y, which is a robust measure of the quality of the field collection and central physician assignment of causes (Aleksandrowicz et al. 2014; Gomes et al. 2017). Ethics approval for the MDS was obtained from St. John’s Research Institute and St. Michael’s Hospital, Toronto, Canada. All SRS participants provided oral consent at the beginning of each sampling frame. Households were free to withdraw, but <1% did so (Jha et al. 2006).

Deaths

We included the entire population who resided in the 2004–2013 sampling units. We focused on deaths from respiratory disease, IHD, stroke (Ke et al. 2018), and all nonaccidental deaths (ICD-10 codes shown in Table 1 footnote) at ages 15–69 y for the main analysis (Table S3). Indian life expectancy in 2018 was 69 y, and we used this age range to consider the younger age distribution of vascular death in India than in high-income countries (Ke et al. 2018). Moreover, the verbal autopsy instrument has lower levels of misclassification before age 70 y than above (Jha et al. 2019). To minimize misclassification of causes, the main analyses was of cause-specific deaths when both physicians initially agreed on the diagnosis, a smaller number than based on adjudication. In India, cardiovascular and respiratory diseases accounted for ~30% of total adult deaths in the period 2010–2013 (RGI and CGHR 2015). Lung cancer deaths are, as documented earlier, uncommon in India (Dikshit et al. 2012; Jha et al. 2008) and were too few to be quantified reliably.

Geocoding and Ambient PM$_{2.5}$ Estimates Linkage

We geocoded 7,416 MDS sampling units using the postal codes in urban areas and village names in rural locations. The SRS records sampling areas as either rural or urban relying on established Census of India’s definitions, based chiefly on population
Table 1. Descriptive statistics of covariates for cause-specific ($n = 56,337$), nonaccidental ($n = 177,686$), and all-cause ($n = 212,573$) deaths, as well as statistics for the general population ($n = 678,889,985$) from India (2004–2013) included in main analysis and death counts for sensitivity analysis.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Stratum</th>
<th>Chronic respiratory</th>
<th>IHD</th>
<th>Stroke</th>
<th>Nonaccidental</th>
<th>All-cause</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death count/population count</td>
<td>—</td>
<td>13,185</td>
<td>29,834</td>
<td>13,118</td>
<td>177,686</td>
<td>212,573</td>
<td>678,889,985</td>
</tr>
<tr>
<td>Age [y (median; 25th, 75th percentile)]</td>
<td>—</td>
<td>62 (56, 66)</td>
<td>60 (53, 65)</td>
<td>56 (44, 64)</td>
<td>54 (39, 63)</td>
<td>27 (12, 42)</td>
<td></td>
</tr>
<tr>
<td>PM$_{2.5}$ [$\mu g/m^3$; median (25th, 75th percentile)]</td>
<td>—</td>
<td>27.1 (18.7, 44.5)</td>
<td>23.5 (16.3, 39.8)</td>
<td>24.4 (16.8, 35.2)</td>
<td>24.4 (16.9, 38.6)</td>
<td>24.3 (16.8, 38.3)</td>
<td>25.5 (17.3, 41.4)</td>
</tr>
<tr>
<td>Percentage of household using solid cooking fuel at subdistrict level$^a$ [median (25th, 75th percentile)]</td>
<td>—</td>
<td>87.1% (67.2%, 94.0%)</td>
<td>72.3% (38.7%, 88.2%)</td>
<td>83.8% (57.0%, 93.4%)</td>
<td>81.5% (54.1%, 93.1%)</td>
<td>81.5% (54.7%, 93.0%)</td>
<td>81.8% (50.7%, 93.4%)</td>
</tr>
<tr>
<td>Percentage of respondent smoked at sampling unit level$^b$ [median (25th, 75th percentile)]</td>
<td>—</td>
<td>11.9% (6.1%, 20.0%)</td>
<td>7.3% (3.2%, 13.3%)</td>
<td>9.8% (4.8%, 17.0%)</td>
<td>9.3% (4.4%, 16.9%)</td>
<td>9.3% (4.4%, 16.7%)</td>
<td>10.0% (4.5%, 18.2%)</td>
</tr>
<tr>
<td>Percentage of female illiteracy at subdistrict level [median (25th, 75th percentile)]</td>
<td>—</td>
<td>57.4% (44.6%, 69.8%)</td>
<td>46.9% (35.3%, 59.7%)</td>
<td>49.5% (37.7%, 61.5%)</td>
<td>51.2% (38.5%, 64.3%)</td>
<td>51.3% (38.8%, 64.3%)</td>
<td>51.0% (38.2%, 64.8%)</td>
</tr>
</tbody>
</table>

| Sex [count (%)] | Female | 5,267 (39.9%) | 8,805 (29.5%) | 5,388 (40.5%) | 71,193 (40.1%) | 82,002 (38.6%) | 331,433,770 (48.8%) |
| Residence [count (%)] | Rural | 10,854 (82.3%) | 20,754 (69.6%) | 9,823 (73.8%) | 133,599 (75.2%) | 161,092 (75.8%) | 499,531,510 (73.6%) |
| | Urban | 2,331 (17.7%) | 9,080 (30.4%) | 3,495 (26.2%) | 44,087 (24.8%) | 51,481 (24.2%) | 179,358,480 (26.4%) |
| Indian region [count (%)] | Northern | 4,894 (37.1%) | 16,426 (55.1%) | 4,936 (37.1%) | 76,201 (42.4%) | 92,730 (42.6%) | 259,234,670 (38.2%) |

Note: Deaths were due to asthma and chronic obstructive pulmonary disease (chronic respiratory) (ICD-10: J40-J47, J63, J95), ischemic heart disease (IHD) (ICD-10: I20-I25, I44, I46, I70, R55, R96), stroke (ICD-10: I60-I69, G45-G46, G51-G53), nonaccidental causes (ICD-10: A00-R99), and all-cause (ICD-10: A00-R99, U00-Y99). Covariate information included in main analyses for ages 15–69 y is presented, and death counts used on stratified sensitivity analyses are summarized. —, no data; PM$_{2.5}$: particulate matter with aerodynamic diameter equal to or smaller than 2.5 μm.

$^a$The percentages of households using solid fuels for cooking (coal, firewood, dung, and crop residue) at the subdistrict-level were obtained from the Census of India 2001 and 2011; the percentages were linked to each death based on the residence locations, with deaths from 2004–5 and 2006–13 used percentages from Census 2001 and 2011, respectively.

$^b$Smoking percentage (prevalence of either bidis or cigarettes) were used as a covariate in Poisson models.
size and population density (RGI and Census Commissioner 2011b). Urban sampling units were geocoded to postal code because postal code information was commonly available for the death records, and rural sampling units to village locations because rural sampling units are located within villages. Only 0.2% of urban sampling units and 4.0% of rural sampling units could not be geocoded due to unavailability of postal code or village information. Figure S4 displays the locations of the geocoded units. We overlaid the geocoded sampling units on the grid cells of the satellite-derived PM2.5 estimates. We performed exploratory analyses on lag years of PM2.5 by linking the 3-y median of ambient PM2.5 estimates to deaths that occurred 2–4 y, 3–5 y, and 4–6 y later based on year of death (e.g., for a 3–5 y lag, 3-y medians of 1999–2001 for PM2.5 were linked to deaths from year 2004; see Figure S5). We found that using lag 2–4 y would cause PM2.5 exposures to be protective on respiratory disease and IHD, whereas using lag 4–6 y the effect of PM2.5 exposures on stroke was smallest in comparison with lag 2–4 y and lag 3–5 y (Figure S6). Thus, to be consistent for all three disease outcomes and avoiding an implausible protective effect of PM2.5, we chose to use lag 3–5 y. Using the 3–5 y lag, the effect of PM2.5 exposures had stronger association with respiratory and IHD mortality, and intermediate association with stroke mortality (Figure S6). We assigned the 3-y median PM2.5 values from the period of 1999–2010 (i.e., a total of ten 3-y intervals) and linked 3-y median of ambient PM2.5 estimates to deaths 3–5 y later in each village or urban block. We were able to link 99.8% of deaths to a satellite-based PM2.5 estimates.

**Statistical Methods**

We modeled each individual’s mortality risk as a function of individual-level risk factors (age and sex), sampling unit-level risk factors (urban/rural, smoking prevalence), a time trend, and spatially attributable risk factors (PM2.5, female illiteracy, dominant language). Further, we allowed for random variation at the unit-level and a smoothly varying spatial random effect. We used model-based generalized linear geostatistical model (Diggle and Ribeiro 2007), currently the state-of-art method for modeling point-referenced spatial data in a hierarchical modeling framework (Gelfand and Banerjee 2017). The approach is broadly similar to the methods used by Bhatt et al. (2015) for estimating malaria prevalence in Africa.

In our model, an individual in age–sex group j living in sampling unit i at time t has a mortality risk $\lambda_{ijt}$, where $\theta$ is the baseline rate for group j in the Indian population, and $\lambda_{ijt}$ is the relative risk (RR). Because the probability of death in any given year is small, we can approximate the binary-valued living/dead response variable with the commonly used Poisson model for the unit-level annual mortality counts (Lawson 2018). The RR by age and sex group are estimated by fitting a simple Poisson regression model with age, sex, and time as explanatory variables; with a large data set the uncertainty in the estimated $\theta$ is negligible. By treating these age–sex “reference rates” as fixed, the observed death counts and age–sex adjusted expected counts for each unit and year are sufficient statistics for making inference on the unit-level and spatial risk factors (see Li et al. 2012). The resulting response variable $Y_{it}$ is the total death count at sampling unit i and time t, which is Poisson distributed with mean $\lambda_{it}P_{it}\theta$, where $P_{it}$ is the vector of population counts by age–sex group j. Our model ignored multiple adult deaths in one household because these are rare (on average, 0.2% within 1 study year) in the MDS. The model is fit separately for each outcome (respiratory disease, IHD, and stroke).

Our models have the following formula for the RR $\lambda_{ijt}$ for individuals living in sampling unit i at time t:

$$Y_{it} \sim \text{Poisson}(\lambda_{ijt}E_{it})$$

$$\log(\lambda_{ijt}) = X_{it}[\beta + W(s_i,t) \gamma + g(PM_{2.5}(s_i,t)) + U(s_i) + Z_i]$$

$$Z_i \sim N(0, \sigma_i^2)$$

$$U(s_i) \sim N(0, \sigma_U^2)$$

$$\text{cor}(U(s + h), U(s)) = \text{Matern}(|h|/\phi; 1).$$

Here $E_{it}$ is the expected number of cases (see Supplemental Material, “Statistical Methods”), $s_i$ denotes the spatial location of sampling unit i, $X_{it}$ are covariates at the unit-level, and spatially referenced covariates are denoted $(s_i, t)$. The satellite-derived PM2.5 value $PM_{2.5}(s_i, t)$ has a nonlinear effect on risk, specified by the function $g(\cdot)$ which can be thought of as a “wiggly” line. More specificaly, $g(\cdot)$ is a second-order random walk with variance parameter $\gamma$, the smaller the value of $\gamma$ the straighter $g(\cdot)$ will be. To be compatible with the most recent global exposure of mortality model (GEMMM) (Burnett et al. 2018), we assumed that PM2.5 exposure in the range of 0–2 $\mu$g/m$^3$ (which is rare in India) had no effect on mortality (values below 2 were discretized to the value of 2). The $X_{it}$ includes an intercept, a linear time trend, urban/rural residency, and smoking prevalence (bidi or cigarette) in the sampling unit (i.e., percentage of smokers among the MDS survey respondents). The spatially referenced covariates $(s_i, t)$ include subdistrict-level (each subdistrict area contains approximately 1–2 MDS sampling units) percentage of female illiteracy (continuous; as a proxy for poverty) and dominant language groups (as a proxy for cultural and dietary differences, which may also be clustered geographically); both were derived from the 2011 Indian census (Figure S7). We did not add subdistrict-level solid fuel use from the 2011 Indian census as a covariate because it contributes to ambient PM2.5.

To account for potential spatial variation, the model included smoothly varying spatial random effects (i.e., spatial autocorrelation or clustering) $U(s_i)$. Similarly, we included independent unit-level random effects $Z_i$ to account for sampling unit-level variation. Unlike the spatial effect $U(s_i)$, two villages in close proximity can have very different values of $Z_i$. Thus, the $Z_i$ can be thought of as accounting for short-scale spatial variation or village-level risk factors not included in the model as covariates, whereas $U(s_i)$ can be interpreted as representing any unknown or unobserved spatially varying risk factors. The spatial random effect $U(s_i)$ has the correlation between the RR at different sampling units locations specified by a Matérn spatial correlation function with a shape parameter fixed at 1.0. The spatial range parameter $\phi$ controls how quickly correlation decays with distance, with large values of $\phi$ leading to smooth surfaces. The variance parameter $\sigma_U^2$ controls the importance that this spatial effect has in determining mortality risk, with $\sigma_U^2$ near zero corresponding to a small effect (see Moraga 2019).

The regression coefficients $\beta$, $\alpha$ were given uninformative normal priors. The prior distributions for $\tau$, $\sigma$, and $1/\phi$ have exponential priors following the “penalized complexity” framework of Simpson et al. (2017). These priors discourage a spatial effect (wanting $U(s_i)$ flat and close to zero) unless the data indicate a clear preference for a spatial model. When the posterior estimate for $\sigma$ is far away from zero and its credible interval is relatively narrow, there is evidence that spatial adjustment is necessary as informed by the data.

We applied two additional variations on our model for the main analysis. First, we treated PM2.5 exposure as a linear effect, estimating RRs for a 10 $\mu$g/m$^3$ increase [with $g(\cdot)$ being linear with a slope parameter]. Second, we ran linear and nonlinear models that did not include adjustment for spatial clustering, that is, we omitted the $U(s_i)$ term in our model (or equivalently set $\sigma = 0$). We compared our results of nonlinear effects of PM2.5 to the GEMM (Burnett et al. 2018) and the Prospective Urban Rural
Epidemiology (PURE) study, which examined nonfatal vascular events and deaths in about 750 communities in 21 countries (Hystad et al. 2020). The cohort studies included in the GEMM adjusted for individual characteristics and contextual variables, but most studies did not include community-level random effects or spatial random effects (Burnett et al. 2018). The PURE study adjusted for individual- and household-level characteristics, geographic covariates, as well as community-level random effects. We used R packages geostatsp (Brown 2015) and R-INLA (Rue et al. 2009) for statistical analyses. We report the median RR or excess risks (defined as RR minus 1) and 95% credible intervals (CI) from the posterior distributions of the Bayesian estimates. Further details of the model fitting and the prior distributions used can be found in Supplemental Material, “Statistical Methods.”

We performed sensitivity analyses focusing on the linear effect of PM$_{2.5}$ with spatial adjustment (by the smoothly varying spatial random effects) to explore whether the RR for PM$_{2.5}$ exposure differed across subgroups using stratified models. Interactions between PM$_{2.5}$ and other potential risk factors were explored by fitting stratified models. We stratified by age (15–44, 45–69, and 70 y and above), by sex, in areas with high (87.3%) and low (0.4%) percentage of households using solid cooking fuels (including firewood, crop residue, dung, and coal) at the subdistrict-level. We obtain solid cooking fuel use percentages at the subdistrict-level from the Census of India 2001 and 2011; the percentages were linked to each death based on the residence locations, with deaths from 2004–2005 and 2006–2013 used percentages from 2001 and 2011 censuses, respectively. We used tertiles to classify each deceased individual into these high, medium, and low solid fuel use levels. Stratification also included rural sampling units (urban sampling units were too few to produce estimate reliably) and northern vs. southern region of India (based on zonal councils https://www.mha.gov.in/zonal-council). Additional sensitivity analyses excluded language group from the covariates and included climatic variables. For climatic variables, we included temperature and relative humidity as covariates in the models. We calculated their annual monthly averages for climatic variables, we included temperature and relative humidity as covariates in the models. We obtained monthly gridded data of temperature and relative humidity from the Center for Weather and Climate Research (CGHR 2015), were substantially lower after spatial adjustment. Further stratification for rural areas, and additional climatic covariates led to only minor changes in the RR. Similarly, the RR for IHD became nonsignificant after spatial adjustment, from 1.02 (95% CI: 1.01, 1.04) to 1.00 (95% CI: 0.97, 1.03) and varying little in rural areas or with additional climatic covariates (Figure 2). By contrast, the RR for stroke was higher with spatial adjustment [1.09 (95% CI: 1.04, 1.14)] than without spatial adjustment [0.91 (95% CI: 0.89, 0.93)], representing a statistically significant excess risk of 9% (Figure 2). For each of the above three conditions, the OR for death compared to injury controls in the case–control analyses showed similar results to the cohort analysis (Figure 2). The association of PM$_{2.5}$ exposure lagged 3–5 y with all nonaccidental deaths varied only slightly without and with adjustment for spatial clustering [RR 0.99 (95% CI: 0.98, 1.00) vs. 1.01 (95% CI: 1.00, 1.03), respectively]. For all-cause mortality, the spatially adjusted RR was 1.02 (95% CI: 1.01, 1.03), and remained statistically significant in the stratified analyses, except exposure to PM$_{2.5}$ rose in India (3-y median of 25 µg/m$^3$ in 1998–2000 to 40 µg/m$^3$ in 2012–2014), with notable hot spots in the north of India, south of the Himalayan mountain range, and West Bengal (Figure 1). The level of PM$_{2.5}$ varies by regions, with central and eastern regions having the highest increase in PM$_{2.5}$ level and the southern region with the least increase (Figure S9).

At ages 15 y or older, we excluded 1.2% of 96,359 respiratory and vascular disease deaths because of missing covariates (Table S4). The excluded population were slightly older, located in sampling units with higher 3-y median of ambient PM$_{2.5}$ satellite-based estimates lagged 3–5 y and had other slight differences from those included (Table 1; Table S4). The main comparison was of 27,553 deaths from chronic respiratory disease, 42,950 from IHD and 24,940 from stroke, or a total of 95,443 deaths, including 56,337 at ages 15–69 y (Table S3; Table S5–7), drawn from a total population of 6.8 million (Table 1). The remaining deaths occurred above age 70 y (Table S8). Table 1 provides summary statistics of the covariates by disease condition. At ages 15–69 y, the median ages at death from chronic respiratory disease, IHD, stroke, all-cause, and nonaccidental causes were older than the median age of the Indian population, and hence all analyses adjusted for age. Deaths from chronic respiratory disease were more rural than in the overall population, had the highest mean PM$_{2.5}$ levels of exposure, more smokers in the sampling unit, higher female illiteracy, and a higher percentage of solid fuel use in the area than in the overall population. IHD deaths occurred more in urban residents than the overall population, and thus these deaths reported lower PM$_{2.5}$ levels of exposure, fewer smokers in the sampling unit, lower solid fuel use, and lower female illiteracy. Stroke deaths showed levels of PM$_{2.5}$ exposure, smoking prevalence at sampling unit level, urban residence, and female illiteracy at subdistrict-level that were generally similar to those of the overall population. The geographic distributions of the three main conditions were distinctive as evident from the distributions by language group regions (Table S5; Figure S7A). More deaths occurred in the northern region, with the exception of IHD, which occurred more in the southern region (Table 1).

After adjusting for death year, urban/rural residency, smoking prevalence in sampling units, subdistrict-level female illiteracy, and dominant language groups, the RRs at ages 15–69 y for both sexes for every 10-µg/m$^3$ increase in 3-y median of ambient PM$_{2.5}$ lagged 3–5 y varied with adjustment for spatial clustering (Figure 2). The RR for chronic respiratory disease deaths, which occurs at much higher rates in north India (RGI and CGHR 2015), were substantially lower after spatial adjustment, from 1.05 (95% CI: 1.03, 1.08) to 1.01 (95% CI: 0.96, 1.05), representing a nonsignificant 1% excess risk. Further stratification for rural areas, and additional climatic covariates led to only minor changes in the RR. Similarly, the RR for IHD became nonsignificant after spatial adjustment, from 1.02 (95% CI: 1.01, 1.04) to 1.00 (95% CI: 0.97, 1.03) and varying little in rural areas or with additional climatic covariates (Figure 2). By contrast, the RR for stroke was higher with spatial adjustment [1.09 (95% CI: 1.04, 1.14)] than without spatial adjustment [0.91 (95% CI: 0.89, 0.93)], representing a statistically significant excess risk of 9% (Figure 2). For each of the above three conditions, the OR for death compared to injury controls in the case–control analyses showed similar results to the cohort analysis (Figure 2). The association of PM$_{2.5}$ exposure lagged 3–5 y with all nonaccidental deaths varied only slightly without and with adjustment for spatial clustering [RR 0.99 (95% CI: 0.98, 1.00) vs. 1.01 (95% CI: 1.00, 1.03), respectively]. For all-cause mortality, the spatially adjusted RR was 1.02 (95% CI: 1.01, 1.03), and remained statistically significant in the stratified analyses, except...
in the southern region (Figure 2). The 2% excess risk in all-cause mortality was mostly driven by stroke mortality, as excluding stroke from all-cause mortality risks yielded a nonsignificant RR of 1.01 (95% CI: 0.99, 1.02). The RRs in the southern region were consistently lower than in the northern region, but with wider credible confidence levels (Figure 2). The RRs excluding language as a covariate were similar to the main results (Figure S10).

Table S9 provides the effects estimates for all covariates and parameters of random effects for models with spatial adjustment. The variance parameters for unit-level (τ) and spatial random effects (σ) have narrow CIs and are far away from zero, showing that both random effects are necessary to be included in the models.

Table 2 presents results stratified by sex, age group, and different levels of solid fuel use. Results for men and women were broadly similar, as were the results for ages 70 y or older. Stratification by high, medium, or low levels of solid fuel use at the local area also yielded broadly similar results, although the RR for chronic respiratory and stroke deaths in men showed higher RR than in women. Sensitivity analysis showed that the
Figure 2. Linear RRs per 10 microgram per cubic meter increase in PM$_{2.5}$ for chronic respiratory disease (n = 13,185), IHD (n = 29,834), stroke (n = 13,318), nonaccidental causes (n = 177,686), and all-cause mortality (n = 212,573) at ages 15–69 y, both sexes combined, from India in 2004–2013, for every 10 micrograms per cubic meter increase in PM$_{2.5}$ with and without spatial adjustment and in various sensitivity analyses. Linear effects for RRs were estimated within the generalized linear model framework, using the Poisson regression model. Models adjusted for death year, urban/rural residency, smoking prevalence in the sampling unit, subdistrict-level percentage of female illiteracy, and dominant language groups (from 2011 Indian census) and included smoothly varying spatial random effects and independent unit-level random effects. Box sizes are weighted by the sample sizes. Horizontal lines represent the 95% CI. Note: CI, credible interval; IHD, ischemic heart disease; OR, odds ratio; PM$_{2.5}$, particulate matter with aerodynamic diameter equal to or smaller than 2.5 μm; RR, relative risk.

* We fitted a logistic model within the generalized linear model framework, adjusting for individual sex, death year, urban/rural residency, age (continuous years), smoking (none, only cigarettes, only bidi, both) interacted with sex, and subdistrict percentage of female illiteracy and language groups. The logistic model used injury deaths as controls, which were shown to have zero association with exposure to PM$_{2.5}$ in a preliminary analysis based on MDS deaths occurring from 2001 to 2003 and PM$_{2.5}$ exposure levels in a special survey done in February 1998 (Figure S8). The reported estimates from the logistic model are OR.
RRs for female deaths were broadly similar when stratified by households with or without male smokers (Figure S11).

The estimated nonlinear models showed that RRs, without adjustment for spatial clustering, at ages 15–69 y in both sexes were higher with increasing levels of PM$_{2.5}$ for chronic respiratory and IHD deaths, but these RRs were lower notably with increasing PM$_{2.5}$ after adjustment for spatial clustering (Figure 3). By contrast, stroke deaths showed lower RR without adjustment but after adjusting for spatial clustering, these RRs were higher with increasing PM$_{2.5}$. Similarly, nonaccidental and all-cause deaths showed higher RR after adjusting for spatial clustering, with increasing PM$_{2.5}$ associated with higher RR. For all three conditions, nonaccidental, and all-cause deaths, RRs at higher levels of exposure had wide credible intervals. The exposure–response relationships for chronic respiratory disease and IHD deaths in the MDS were much flatter than the modeled relationship of hazard ratios from the GEMM but were more comparable for stroke deaths (Figure 4). The exposure–response relationships of PM$_{2.5}$ with mortality in the MDS were similar to those of incident IHD or stroke in the comparable analyses used by the PURE study (Hystad et al. 2020; Figure 4). GEMM showed the steepest relationship at the youngest adult ages, whereas the MDS showed no such differences by age (Table 2). Even the 95% upper CI of the estimates showing RRs of 1.05, 1.03 and 1.14, respectively, for chronic respiratory disease, IHD, or stroke (Figure 2), were considerably lower than the mean GEMM modeled estimates (Burnett et al. 2018).

Discussion

To our knowledge, our large retrospective cohort study provides the first nationally representative epidemiological evidence of the association between cause-specific mortality and long-term PM$_{2.5}$ exposure based on lagged satellite-derived PM$_{2.5}$ exposure. We find a statistically significant excess risk of 9% (range 4%–14%) for stroke deaths at ages 15–69 y per 10-μg/m$^3$ increase in exposure, but no significant excess risks for chronic respiratory disease, IHD deaths, or all nonaccidental deaths. The overall mortality excess risk was 2% (range 1%–3%) but was nonsignificant once stroke was excluded. The RRs included adjustment for both spatial and unit-level variation in risks unexplained by covariates included in the models and were consistent in various sensitivity analyses including stratifying by age, sex, levels of solid fuel use, in rural areas, and focusing only on the northern regions where PM$_{2.5}$ exposure is most widespread. Results on case–control analyses using injury controls showed nearly identical results once spatial clustering was considered.

The MDS association of PM$_{2.5}$ exposure with IHD deaths is consistent with a zero excess risk. Respiratory disease deaths also showed close to a zero excess risk, although we noted higher risks among men than women, perhaps reflecting the residual effects of smoking history not captured in the community measures of smoking prevalence (given that few women smoke in India). The story for stroke is more nuanced and even our observed 9% excess risk may not be real. Stroke mortality is lowest in several northern states adjacent to the Himalayas: Punjab, Haryana, Rajasthan, Uttar Pradesh, and Bihar (Ke et al. 2018), states which coincidentally have the highest levels of satellite-measured ambient PM$_{2.5}$ estimates (Figure 1). In the analysis without spatial adjustment, therefore, PM$_{2.5}$ appears to be protective against stroke, an artifact that reduces with spatial adjustment. The model with spatial adjustment looked for a consistent relationship between more regional variation in PM$_{2.5}$ and stroke mortality. Once we considered a smooth spatial effect to explain the large region of low stroke mortality in the north, we identified a modest but significant excess risk.

The RRs for adult respiratory disease mortality are modest in comparison with published model-based estimates. Chronic respiratory disease is more prevalent in the northern states, Rajasthan, Uttar Pradesh, and Maharashtra (Salvi et al. 2018). For IHD, higher prevalence is found in Punjab in the north and in Andhra Pradesh and Tamil Nadu in the south (Menon et al. 2019). Thus, deaths from chronic respiratory diseases (and to a lesser extent IHD) cluster in the same areas as does PM$_{2.5}$ exposure, even after taking into account differences in subdistrict-level illiteracy rate or sampling unit–level smoking prevalence. The PURE study, which included community-level random effects in their models, also showed a lower RR for nonvascular deaths. Our estimates for stroke are comparable to the PURE results (Hystad et al. 2020).

The exposure variable is 3-y median of ambient and satellite-derived PM$_{2.5}$ estimates, which is widely used in environmental epidemiology due to its availability. At a minimum, our results show that the large RRs reported for satellite-derived PM$_{2.5}$ elsewhere are not reproducible in India. However, the MDS results are not directly comparable to the modeling results from GEMM or the Global Burden of Disease (GBD) for several reasons. First, although our IHD RRs are comparable to the RRs in the PURE study and in some of the larger studies that were inputs to the GEMM/GBD meta-analyses, the modeled GEMM/GBD results comprise a meta-analysis of 41 cohort studies, with various additional transformations to create exposure-hazard relationships. Unfortunately, most of the input studies did not consider spatial clustering (Burnett et al. 2018). In contrast, the MDS estimates did not need to rely on any prior assumptions about exposure–response relationships or transformations of PM$_{2.5}$ estimates, yielding markedly divergent exposure–response curves for IHD and chronic respiratory disease but more similar curves for stroke. Second, the range of 3-y median of ambient PM$_{2.5}$ estimates studied in the MDS is far larger (and captures much higher ambient PM$_{2.5}$ levels) than the relatively narrow range in the GEMM input studies. Last, the MDS RRs for the key conditions are smaller in comparison with GEMM HRs—0% to 9% excess mortality risks with a 10-μg/m$^3$ increase in ambient PM$_{2.5}$ exposure. Even for stroke, we did not observe the near 2-fold hazards that GEMM reported for ambient PM$_{2.5}$ exposure in highly polluted areas (like Delhi and Varanasi), which are close to the excess mortality documented for stroke from bidi or cigarette smoking in India (Jha et al. 2008). Modest risks from ecological exposures are difficult to quantify reliably. Hence, suitable caution is warranted in interpreting the MDS evidence, and even more caution is warranted in extrapolating GEMM/GBD’s current model-based associations to India.

Our findings, if true, suggest a need for a substantial downward revision of the WHO estimates (which effectively use the GBD or GEMM results), which suggest that ambient PM$_{2.5}$ accounts for 0.29 million, 0.39 million, and 0.14 million deaths, respectively, from chronic respiratory diseases, IHD, and stroke at ages 25 y and above in India (WHO 2018). These model-based estimates find that IHD contribute more than one-quarter of the total mortality due to exposure to ambient air pollution in India and worldwide (WHO 2018). However, studies linking ambient PM$_{2.5}$ exposure to IHD have been mostly done in high-income countries where the levels of PM$_{2.5}$ concentrations are lower and the range of exposure narrower (Burnett et al. 2018; Kaufman et al. 2016). We find a negligible association for IHD with PM$_{2.5}$ exposure in India. This finding might reflect differences in the susceptibility to PM$_{2.5}$ in the Indian population and/or a difference in the composition and resulting toxicity of PM$_{2.5}$ exposure in India. On the exposure side, we cannot exclude possibly missing more refined PM$_{2.5}$ exposure that could increase vascular risk, such as ultrafine particles that may impact vascular end points (Schulz et al. 2005) and that may differ across countries.
Figure 3. Nonparametric RR per unit (micrograms per cubic meter) increase in PM$_{2.5}$ exposure for chronic respiratory disease ($n = 13,185$), IHD ($n = 29,834$), stroke ($n = 13,318$), nonaccidental causes ($n = 177,686$), and all-cause mortality ($n = 212,573$) at ages 15–69 y, both sexes combined, from India in 2004–2013, with and without spatial adjustment. Nonparametric effects for RRs were estimated using the Poisson regression model. Models adjusted for death year, urban/rural residency, smoking prevalence in the sampling unit, subdistrict-level percentage of female illiteracy, and dominant language groups (from 2011 Indian census) and included smoothly varying spatial random effects and independent unit-level random effects. Solid curves are the median RR estimates; dotted curves are 95% CI (graph is limited to RRs between 0.0 and 2.5). Vertical black arrows at x-axis indicate national population-weighted average exposure of 47 µg/m$^3$ ambient PM$_{2.5}$ in 2014. Note: CI, credible interval; IHD, ischemic heart disease; RR, relative risk.
Chronic respiratory

Relative Risk / Hazard Ratio

Satellite PM$_{2.5}$ ($\mu g/m^3$)

GEMM: COPD
MDS: 15–69 yr

HID

Relative Risk / Hazard Ratio

Satellite PM$_{2.5}$ ($\mu g/m^3$)

GEMM: 50–54 yr
GEMM: 70–74 yr
PURE: 35–70 yr
MDS: 15–69 yr

Stroke

Relative Risk / Hazard Ratio

Satellite PM$_{2.5}$ ($\mu g/m^3$)

GEMM: 50–54 yr
GEMM: 70–74 yr
PURE: 35–70 yr
MDS: 15–69 yr

Figure 4. Comparison of exposure–response relationships of PM$_{2.5}$ risk for chronic respiratory disease (top panel, n = 13,185), IHD (middle panel, n = 29,834), and stroke (lowest panel, n = 13,318) deaths in GEMM, PURE, and MDS from India, 2004–2013. Estimates are at various ages in GEMM (based on hazard ratios) and incidence of IHD and strokes in PURE at ages 35–70 y (based on hazard ratios) and in the MDS at ages 15–69 y (based on RRs). GEMM composed of a meta-analysis of 41 cohort studies with various transformations to create exposure–hazard relationships. Estimates for COPD from GEMM is not age specific, MDS estimates were adjusted for death year, urban/rural residency, smoking prevalence in the sampling unit, subdistrict-level percentage of female illiteracy, and dominant language groups (from 2011 Indian census) and included smoothly varying spatial random effects and independent unit-level random effects. PURE estimates were adjusted for age, sex, baseline year, smoking status, physical activity, PURE diet index score, waist-to-hip ratio, INTERHEART risk score, use of solid fuels for cooking, education level, household wealth index, occupational class, baseline cardiovascular disease and chronic conditions, cardiovascular disease medication use, hypertension, geographical covariates (urban or rural location, baseline country gross domestic product per capita, Night Light Development Index score, and a national or regional Health Care Access and Quality Index), and community random effect (Hystad et al. 2020). The comparison is to PURE model 3, which the authors of the study indicate is their preferred model. For GEMM’s estimates for IHD and stroke, younger ages have higher hazard ratios (or steeper exposure–response curves) than older ages. Shaded color band represent regions of 95% CI for MDS estimates. Note: CI, credible interval; COPD, chronic obstructive pulmonary disease; GEMM, global exposure of mortality model; IHD, ischemic heart disease; MDS, Million Death Study; PURE, Prospective Urban Rural Epidemiology.

Figure 4. (Continued.)

(Burney and Amaral 2019) but are not well captured in satellite-based estimates. Notably, the PURE study found a 3% significant excess risk for vascular disease deaths with every 10-µg/m$^3$ increase in PM$_{2.5}$ exposure, but this excess risk disappeared once adjustment was made for dust and salt in satellite-based ambient PM$_{2.5}$ estimates (Hystad et al. 2020). Thus, the effect of PM$_{2.5}$ exposure is highly sensitivity to the type of exposure data and spatial adjustment. More speculatively, low levels of PM$_{2.5}$ exposure, which are uncommon in India, might play a disproportionate role in increasing IHD (Papadogeorgou et al. 2019). Moreover, most IHD deaths in India involve conventional risk factors such as smoking (Jha et al. 2008), elevated blood pressure and blood lipids, and, unusually, low body mass index (BMI) (Gajalakshmi et al. 2018; Ke et al. 2021). Although our estimates of IHD mortality from ambient PM$_{2.5}$ adjusted for smoking prevalence in sampling units, it is possible that smoking type and other risk factors measured at the individual level, such as BMI, diet, alcohol drinking, education level, and medical history, might have confounding effect on the mortality–exposure associations, as might access to effective treatments (Ke et al. 2018). In Chinese populations adjacent to northeast India, the relationship of ambient PM$_{2.5}$ with stroke is largely with hemorrhagic strokes (Chen et al. 2018). Because verbal autopsy cannot distinguish between hemorrhagic and occlusive strokes, it is possible that the notably high stroke mortality in northeastern India noted earlier (Ke et al. 2018) might represent a different relationship than that found in China.

Several limitations exist in the MDS study, including measurement error in the ambient PM$_{2.5}$ estimates, the potential omission of additional confounders (i.e., solid fuel use), and possible misclassification of the mortality outcomes. The key consideration is whether these various forms of model misspecification have attenuated our estimated RRs toward one, meaning we were unable to detect modest but perhaps clinically relevant excess risks. Overall, we do not believe we have systematically underestimated the RRs. First, adjustment for spatial clustering is appropriate for dealing with residual correlation in spatial data (Dormann et al. 2007). Cohort analyses for the United States have generally included random effects to represent spatial patterns in the data and found that hazards were significantly impacted by their inclusion (Krewski et al. 2009). Second, the uncertainties that exist in our PM$_{2.5}$ exposure assessment method could cause either overestimation or underestimation of RRs (van...
Table 2. Stratification analyses of RRs per micrograms per cubic meter increase in PM$_{2.5}$ exposure, by age group and solid fuel use level for male and female deaths from chronic respiratory disease ($n$ = 27,504), IHD ($n$ = 42,841), stroke ($n$ = 24,899), nonaccidental causes ($n$ = 315,457), and all causes ($n$ = 358,662) in India, 2004–2013.

<table>
<thead>
<tr>
<th>Group</th>
<th>Strata</th>
<th>Chronic respiratory</th>
<th>IHD</th>
<th>Stroke</th>
<th>Nonaccidental causes</th>
<th>All-cause</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases</td>
<td>RR (95% CI)</td>
<td>Cases</td>
<td>RR (95% CI)</td>
<td>Cases</td>
<td>RR (95% CI)</td>
</tr>
<tr>
<td>Male Age group (y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15–49</td>
<td>921</td>
<td>1.08 (0.98, 1.19)</td>
<td>6,628</td>
<td>0.97 (0.93, 1.02)</td>
<td>1,498</td>
<td>1.07 (0.97, 1.18)</td>
</tr>
<tr>
<td>50–69</td>
<td>6,997</td>
<td>1.03 (0.98, 1.09)</td>
<td>14,401</td>
<td>1.02 (0.98, 1.05)</td>
<td>6,432</td>
<td>1.05 (0.99, 1.12)</td>
</tr>
<tr>
<td>70+</td>
<td>8,288</td>
<td>0.97 (0.92, 1.01)</td>
<td>8,016</td>
<td>1.03 (0.98, 1.08)</td>
<td>5,953</td>
<td>1.05 (0.99, 1.11)</td>
</tr>
<tr>
<td>Solid fuel use level*</td>
<td>Low</td>
<td>1,149</td>
<td>1.08 (0.98, 1.18)</td>
<td>5,938</td>
<td>1.00 (0.96, 1.05)</td>
<td>1,559</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>2,897</td>
<td>1.02 (0.95, 1.10)</td>
<td>9,656</td>
<td>1.02 (0.97, 1.07)</td>
<td>3,012</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>3,871</td>
<td>1.07 (1.00, 1.14)</td>
<td>5,430</td>
<td>0.99 (0.94, 1.05)</td>
<td>3,358</td>
</tr>
<tr>
<td>Female Age group (y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15–49</td>
<td>673</td>
<td>0.94 (0.86, 1.03)</td>
<td>2,255</td>
<td>1.03 (0.98, 1.07)</td>
<td>938</td>
<td>0.91 (0.83, 1.01)</td>
</tr>
<tr>
<td>50–69</td>
<td>4,594</td>
<td>1.01 (0.95, 1.07)</td>
<td>6,550</td>
<td>1.03 (0.98, 1.08)</td>
<td>4,450</td>
<td>1.04 (0.97, 1.12)</td>
</tr>
<tr>
<td>70+</td>
<td>6,031</td>
<td>1.04 (0.99, 1.10)</td>
<td>4,991</td>
<td>1.06 (1.00, 1.13)</td>
<td>5,628</td>
<td>1.09 (1.03, 1.17)</td>
</tr>
<tr>
<td>Solid fuel use level*</td>
<td>Low</td>
<td>767</td>
<td>0.98 (0.89, 1.07)</td>
<td>2,432</td>
<td>1.01 (0.96, 1.05)</td>
<td>987</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>1,882</td>
<td>0.98 (0.90, 1.07)</td>
<td>3,860</td>
<td>1.04 (0.98, 1.11)</td>
<td>1,987</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>2,617</td>
<td>1.01 (0.94, 1.09)</td>
<td>2,512</td>
<td>0.99 (0.92, 1.06)</td>
<td>2,413</td>
</tr>
</tbody>
</table>

Note: ICD-10 codes for chronic respiratory are J40-J47, J63, J93; for IHD are I20-I25, I44, I46, I70, R55, R96; for stroke are I60-I69, G45-G46, G81-G83; for nonaccidental causes are A00-R99; and for all-cause mortality are A00-R99, U00-Y99. Estimates were derived from a model-based generalized linear geostatistical model with a Poisson distribution. All models were adjusted for death year, urban/rural residency, sampling unit-level smoking prevalence, subdistrict-level female illiteracy, and dominant language groups and included smoothly varying spatial random effects and independent unit-level random effects. Credible intervals (CI) of 95% were reported for the RR. CI, credible interval; IHD, ischemic heart disease; RR, relative risk.

*The tertiles of the subdistrict-level percentage of households using solid cooking fuels were used to classify each deceased individual into high, medium, and low solid fuel use categories. Data range and median values for each category are provided in the parentheses.
S11). Women have higher levels of exposure to solid fuel use than men (Table 2) and in the model assessing confounding between those who died and the overall population (Table 1), age of the population using solid fuel was broadly similar between those who died and the overall population (Table 1). Women have higher levels of exposure to solid fuel use than men (Choudhuri and Desai 2020), so the broad consistency of the RRs for PM2.5 exposure by sex also suggest that our RRs are not likely to be greatly confounded by solid fuel use. Fourth, there are inherent uncertainties of verbal autopsy assignment of cause of death, even though the distinction between stroke and IHD deaths is reasonably clear with more misclassification between IHD and respiratory disease deaths (Jha et al. 2019). We addressed this uncertainty by relying on dual physicians’ initial agreement on the deceased person’s cause of death, which was independent of PM2.5 level (Table S2). Nevertheless, there remain residual uncertainties in the cause-of-death assignment even using these strict criteria. Use of this strict criteria may also lead to capturing fewer number of cause-specific deaths. Furthermore, such misclassification of causes is greater at ages 70 y or more, but we observed similar RRs in this age group and in the younger age group for the three causes. Our main results focus on ages 15–69 y, where the number of ill-defined causes of death were low in comparison with older ages, to ensure higher accuracy in the reported estimates. Moreover, the use of all nonaccidental deaths yielded similar results to those for IHD and respiratory disease.

Fifth, our analyses have some limitations which are countered by the fact that the sample was nationally representative and drawn from over 7,400 discrete areas and that our models considered both unit-level and spatially varying unmeasured risk factors. We had fewer individual-level covariates than did the PURE study, but our results were very similar. The PURE study also showed little variation in RRs by adding more individual covariates (Hystad et al. 2020). Use of community-level variables potentially introduces residual confounding in the models. We minimized the risk of residual confounding by adjusting for deprivation (via female illiteracy), smoking prevalence in sampling units, various sociocultural risk factors by including language, and, most important, by the adjustment for unit-level variation and spatial clustering. Moreover, we saw nearly identical results from case-control analyses, where any bias affecting assessment of the ambient PM2.5 exposure in the cases should similarly affect assessment of the exposure in control deaths. Our model does not produce hazard ratios typically used in prospective studies. However, our model is equivalent to a survival model, assuming (reasonably) that the hazard function is piecewise constant changing every 5 y of age. Because we are modeling a rare event, with nearly all the death dates being right censored, a survival model with a spatial random effect would be challenging to fit to a data set of this size. Last, we do not report on childhood deaths, including childhood pneumonia, which is the leading cause of childhood mortality in India (MDS Collaborators 2017) and has been linked to PM2.5 exposure (WHO 2018). However, the overall age-standardized death rates from pneumonia at ages 0–4 y and 5–14 y have fallen substantially in India from 2000 to 2015, including in the states where PM2.5 exposure is most common (MDS Collaborators 2017; Fadel et al. 2019).

Given these uncertainties in our estimates, we did not derive the population attributable risks, because we believe it is premature to do so until direct epidemiological studies can further quantify the relationship between ambient PM2.5 exposure and mortality and establish causality (which model-based estimates presuppose). Naturally, errors in measuring PM2.5 apply also to GEMM/WHO estimates, which also rely on these satellite-based estimates. Indeed, our study emphasizes the need for additional reliable nationwide mortality studies that incorporate ambient PM2.5, smoking, and other relevant risk factors in countries with relatively high PM2.5 exposure. Better ground-based measurement to capture the spatiotemporal variations of PM2.5 exposures (such as daily exposure near each residence) is warranted (Kaufman et al. 2016). Our population model, ideally with individual-level smoking type and other confounders and including examining possible interactions between sex, age group, and solid fuel use, could help to resolve some of the uncertainties in documenting mortality from ambient PM2.5 exposure in China (Yin et al. 2017) and elsewhere. However, action to reduce PM2.5 exposure is justified regardless of these uncertainties. There is substantial earlier evidence and biological plausibility that chronic respiratory morbidity and mortality in children from acute respiratory infection can be reasonably attributable to PM2.5 exposure (Gehring et al. 2013). Limiting air pollutant output from traffic, industrial, and household sources is a relevant strategy in India to reduce overall PM2.5 exposure (Karahalios et al. 2015).

Acknowledgments

P.J. conceived the study. R.G., G.D.S., P.B., and P.J. were the study principal investigators. The MDS Collaborators implemented the data collection procedures. P.B., Y.I., P.R., S.H.F., and P.J. conducted the analysis. P.J., P.B., G.D.S., and S.H.F. wrote the first draft, and all authors interpreted the data, reviewed results, and contributed to the report. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. P.J. is the guarantor. The authors thank P. Hystad and S. Yusuf for providing data from the PURE study and H. Gelband for useful comments. This study was supported by the Indian Council of Medical Research, Canadian Institutes of Health Research Foundation grant (FDN 154277), the University of Toronto, the National Institutes of Health grants (R01TW007939). This paper expresses the views of the authors and working group and does not necessarily represent the views of the Government of India. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing the report.

Data availability: Please refer to https://censusindia.gov.in/censuswebsite/data/SRSSTAT for public reports. For MDS data access procedures, please contact the Office of the Registrar General, RK Puram, New Delhi, India. The satellite-derived ambient PM2.5 data from van Donkelaar et al. 2015 can be obtained free of charge from https://sites.wustl.edu/acag/datasets/surface-pm2-5/. The 2011 Indian census data (including data on female illiteracy, language groups, and religion) can be accessed from https://data.gov.in. The Gridded Population of the World (version 4) can be downloaded from

Environmental Health Perspectives
097004-12 130(9) September 2022