Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter

Richard Burnett[a], Hong Chen[a,b], Mieczysław Szyszkowicz[a,1], Neal Fann[c], Bryan Hubbell[d], C. Arden Pope III[e], Joshua S. Apte[f], Michael Brauer[g], Aaron Cohen[h], Scott Weichenthal[i,j], Jay Coggins[k], Qian Di[l], Bert Brunekreef[m], Joseph Frostad[n], Stephen S. Lim[n], Haidong Kan[o], Katherine D. Walker[h], George D. Thurston[p], Richard B. Hayes[q], Chris C. Lim[r], Michelle C. Turner[s], Michael Jerrett[t], Daniel Krewski[u], Susan M. Gapstur[v], W. Ryan Diver[v], Bart Ostro[w], Debbie Goldberg[x], Daniel L. Crouse[y], Randall V. Martin[z], Paul Peters[aa,bb,cc], Lauren Pinault[dd], Michael Tjepkema[dd], Aaron van Donkelaar[z], Paul J. Villeneuve[aa], Anthony B. Miller[ee], Peng Yin[ff], Maigeng Zhou[ff], Lijun Wang[ff], Nicole A. H. Janssen[gg], Marten Marra[gg], Richard W. Atkinson[hh,ii], Hilda Tsang[jj], Thuan Quoc Thach[jj], John B. Cannon[e], Ryan T. Allen[e], Jaime E. Hart[kk], Francine Laden[kk], Giulia Cesaroni[ll], Francesco Forastiere[ll], Gudrun Weinmayr[mm], Andrea Jaensch[mm], Gabriele Nagel[mm], Hans Concin[nn], and Joseph V. Spadaro[oo]

[a]Population Studies Division, Health Canada, Ottawa, ON K1A 0K9, Canada; [b]Department of Environmental and Occupational Health, Public Health Ontario, Toronto, ON M5G 1V2, Canada; [c]Risk and Benefits Group, Office of Air Quality Planning and Standards, US Environmental Protection Agency, Washington, DC 20460; [d]Office of Research and Development, US Environmental Protection Agency, Washington, DC 20460; [e]Department of Economics, Brigham Young University, Provo, UT 84602; [f]Department of Civil, Architectural and Environmental Engineering, University of Texas at Austin, Austin, TX 78712; [g]School of Population and Public Health, University of British Columbia, Vancouver, BC V6T 1Z3, Canada; [h]Health Effects Institute, Boston, MA 02110-1817; [i]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC H3A 0G4, Canada; [j]Gerald Bronfman Department of Oncology, McGill University, Montreal, QC H3A 0G4, Canada; [k]Department of Applied Economics, University of Minnesota, Minneapolis, MN 55455; [l]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA 02115; [m]Institute for Risk Assessment Sciences, Universiteit Utrecht, 3512 JE Utrecht, The Netherlands; [n]Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA 98195; [o]School of Public Health, Fudan University, Shanghai 200433, China; [p]Environmental Medicine and Population Health, Program in Human Exposures and Health Effects, New York University School of Medicine, New York, NY 10016; [q]Department of Population Health, NYU Langone Medical Center, New York, NY 10016; [r]Department of Environmental Medicine, New York University School of Medicine, New York, NY 10016; [s]ISGlobal, Barcelona Institute for Global Health, 08036 Barcelona, Spain; [t]Department of Environmental Health Sciences, Fielding School of Public Health, University of California, Los Angeles, CA 90095; [u]McLaughlin Centre for Population Health Risk Assessment, University of Ottawa, Ottawa, ON K1N 6N5, Canada; [v]Epidemiology Research Program, American Cancer Society, Inc., Atlanta, GA 30303; [w]Department of Civil and Environmental Engineering, University of California, Davis, CA 95616; [x]Cancer Prevention Institute of California, Fremont, CA 94538; [y]Department of Sociology, University of New Brunswick, Fredericton, NB E3B 5A3, Canada; [z]Department of Physics and Atmospheric Science, Dalhousie University, Halifax, NS B3H 4R2, Canada; [aa]Department of Health Sciences, Carleton University, Ottawa, ON K1S 5B6, Canada; [bb]Department of Geography and Environment, Carleton University, Ottawa, ON K1S 5B6, Canada; [cc]New Brunswick Institute for Research, Data and Training, University of New Brunswick, Fredericton, NB E3B 5A3, Canada; [dd]Health Analysis Division, Statistics Canada, Ottawa, ON K1A 0T6, Canada; [ee]Dalla Lana School of Public Health, University of Toronto, Toronto, ON M5T 3M7, Canada; [ff]National Center for Chronic Noncommunicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 100050, China; [gg]National Institute for Public Health and the Environment, 3720 BA Bilthoven, The Netherlands; [hh]Population Health Research Institute, St. George's, University of London, London SW17 0RE, United Kingdom; [ii]MRC-PHE Centre for Environment and Health, St. George's, University of London, London SW17 0RE, United Kingdom; [jj]School of Public Health, University of Hong Kong, Hong Kong, China; [kk]Department of Environmental Health, Harvard C.T. Channing School of Public Health, Harvard University, Boston, MA 02115; [ll]Department of Epidemiology, Regional Health Service, ASL Roma 1, 00147 Rome, Italy; [mm]Institute of Epidemiology and Medical Biometry, Ulm University, 89081 Ulm, Germany; [nn]Agency for Preventive and Social Medicine, 6900 Bregenz, Austria; and [oo]Spadaro Environmental Research Consultants (SERC), Philadelphia, PA 19142

Exposure to ambient fine particulate matter (PM$_{2.5}$) is a major global health concern. Quantitative estimates of attributable mortality are based on disease-specific hazard ratio models that incorporate risk information from multiple PM$_{2.5}$ sources (outdoor and indoor air pollution from use of solid fuels and secondhand and active smoking), requiring assumptions about equivalent exposure and toxicity. We relax these contentious assumptions by constructing a PM$_{2.5}$-mortality hazard ratio function based only on cohort studies of outdoor air pollution that covers the global exposure range. We modeled the shape of the association between PM$_{2.5}$ and nonaccidental mortality using data from 41 cohorts from 16 countries—the Global Exposure Mortality Model (GEMM). We then constructed GEMMs for five specific causes of death examined by the global burden of disease (GBD). The GEMM predicts 8.9 million [95% confidence interval (CI): 7.5–10.3] deaths in 2015, a figure 30% larger than that predicted by the sum of deaths among the five specific causes (6.9; 95% CI: 4.9–8.5) and 120% larger than the risk function used in the GBD (4.0; 95% CI: 3.3–4.8). Differences between the GEMM and GBD risk functions are larger for a 20% reduction in concentrations, with the GEMM predicting 220% higher excess deaths. These results suggest that PM$_{2.5}$ exposure may be related to additional causes of death than the five considered by the GBD and that incorporation of risk information from other, nonoutdoor, particle sources leads to underestimation of disease burden, especially at higher concentrations.

mortality | exposure | risk | concentration | fine particulate matter

Exposure to outdoor fine particulate matter (PM$_{2.5}$) is recognized as a major global health concern (1). In particular, both nonaccidental and cause-specific mortality have been associated with outdoor PM$_{2.5}$ concentrations. In cohort studies, where subjects provide information on major mortality risk factors such as cigarette smoking, obesity, and occupation, estimates of outdoor PM$_{2.5}$ exposure are assigned based on multiple year averages and followed over time to ascertain their date and underlying cause of death. The magnitude of the association between PM$_{2.5}$ exposure and the probability of death is described by the hazard ratio (2).

Significance

Exposure to outdoor concentrations of fine particulate matter is considered a leading global health concern, largely based on estimates of excess deaths using information integrating exposure and risk from several particle sources (outdoor and indoor air pollution and passive/active smoking). Such integration requires strong assumptions about equal toxicity per total inhaled dose. We relax these assumptions to build risk models examining exposure and risk information restricted to cohort studies of outdoor air pollution, now covering much of the global concentration range. Our estimates are severalfold larger than previous calculations, suggesting that outdoor particulate air pollution is an even more important population health risk factor than previously thought.

However, the specific shape of this association has not been identified, neither for relatively low exposures in developed Western countries nor higher exposures observed globally.

Until recently, cohort studies of outdoor $PM_{2.5}$ and mortality were limited to areas with relatively low concentrations (<35 µg/m^3) compared with the entire global exposure range (3). This lack of direct evidence at higher global $PM_{2.5}$ concentrations motivated the Integrated Exposure-Response model (IER) (4), which combined information on $PM_{2.5}$–mortality associations from nonoutdoor $PM_{2.5}$ sources, including secondhand smoke, household air pollution from use of solid fuels, and active smoking. Specifically, to construct the IER, estimates of the total mass of inhaled particles from each nonoutdoor source were converted into the equivalent concentration in the ambient atmosphere. The IER forms the basis of the estimates of disease burden attributable to $PM_{2.5}$ (e.g., 4 million deaths in 2015) in the global burden of disease (GBD) (1), those of the World Health Organization (WHO) (5), and in the quantification of impacts of policy scenarios on projected improvements in population health burden and evaluation of air-quality standards (6, 7).

By using this approach, stable predictions of the hazard ratio function can be obtained over the entire global range of outdoor $PM_{2.5}$; however, the IER requires risk information on sources other than outdoor $PM_{2.5}$ and assumes equal toxicity per unit dose across these nonoutdoor sources. Risk assessments of outdoor particles have assumed that toxicity is a function of mass concentration alone (8, 9). The IER extended this assumption to particle sources mainly originating from indoor sources, such as secondhand smoking and heating/cooking, and to particle exposure from active smoking. In addition, the IER assumes that the dosing rate from cigarette smoking, a large intake of particles over repeated short time periods per day, results in the same toxicity as continually breathing the same total dose from the atmosphere per day. Similar assumptions are required for exposure to secondhand smoke and household pollution. For example, the total particle dose from smoking a single cigarette is assumed equivalent to breathing an ambient atmosphere of 667 µg/m^3 for 24 h (4).

The IER formulation also assumes a counterfactual uncertainty distribution, where the relative risk of morality at any concentration is compared with the counterfactual concentration. The uncertainty distribution is defined as a uniform random variable with lower and upper bounds specified by the average of the minimum (2.4 µg/m^3) and fifth percentile (5.9 µg/m^3) concentrations of cohort studies where subjects are exposed to relatively low values (3). This definition was adopted due to lack of knowledge about the shape of the concentration–mortality association at these lower levels.

We seek to relax many of the strong assumptions required by the IER by relying solely on studies of outdoor $PM_{2.5}$. First, we established collaborations between 15 research groups globally that have examined the relationship between long-term exposure to outdoor $PM_{2.5}$ and mortality (10–24). Each of these 15 research groups independently conducted analyses to characterize the shapes of $PM_{2.5}$–mortality associations in their respective cohorts using a hazard ratio function developed for health impact assessment (25). Among these 15 cohorts is a study of Chinese men (10) with long-term outdoor $PM_{2.5}$ exposures up to 84 µg/m^3, thus greatly extending the range of exposures observed in cohort studies conducted in high-income countries in Europe and North America. In 2015, 97% of the global population lived in countries whose population-weighted outdoor exposure was <84 µg/m^3. Our within-cohort analysis focused on nonaccidental mortality as this outcome represents the total mortality burden of $PM_{2.5}$ exposure and provides enhanced statistical power to characterize the shape of the $PM_{2.5}$–mortality associations compared with any specific cause of death. Almost all nonaccidental deaths were due to noncommunicable diseases (NCDs) and lower respiratory infections (LRIs). We thus have restricted our global estimates of excess deaths to this subgroup of illnesses. We were also able to relax the need to assume a counterfactual uncertainty distribution by directly examining the shape of the concentration–mortality association at relatively low levels included in several cohorts.

To complement information from these 15 cohorts, we also extracted data from the published literature (i.e., hazard ratios between $PM_{2.5}$ and nonaccidental mortality) for an additional 26 cohorts where we did not have access to the subject level information (24, 26–33). For the 15 within-cohort analyses, we relaxed the assumption that concentration–mortality associations were linear within each cohort. A linear association between exposure and the logarithm of the baseline hazard ratio was assumed for the remaining 26 cohorts. We then estimated the Global Exposure Mortality Model (GEMM) as a common (possibly nonlinear) hazard ratio model among the 41 cohorts by pooling predictions of the hazard ratio among cohorts over their range of exposure (*SI Appendix, SI Methods*), denoted as GEMM NCD+LRI.

For comparison with previous disease burden assessments, we also constructed separate GEMMs for each of the five causes of death that comprise the GBD attributable mortality estimates: ischemic heart disease (IHD), stroke, chronic obstructive pulmonary disease (COPD), lung cancer, and LRIs. The hazard ratio and exposure information used by GBD2015 for outdoor air pollution (3) was complemented with data on specific causes of death, hazard ratios, and exposure from those studies where we used the nonaccidental risk information but were not included in the GBD2015 IER models. Four studies (11, 13, 17, 19) were included for LRI, three studies (10, 17, 31) for both IHD and stroke, and two studies (10, 17) for COPD and lung cancer. We assumed that the association between $PM_{2.5}$ and the logarithm of the baseline hazard was linear within each cohort in a manner similar to that used by the IER (4).

We note that almost all ($>99\%$) nonaccidental deaths in the 41 cohorts were due to noncommunicable diseases and LRIs (NCD+LRI). Adult (>25 y) mortality rates based on this subgroup are used when calculating excess mortality attributable to $PM_{2.5}$ exposure for the nonaccidental GEMM. Restriction to NCD+LRI causes of death also addresses the issue that some countries have a much higher proportion of communicable disease mortality compared with the 41 cohorts. We now denote the nonaccidental GEMM as GEMM NCD+LRI and the GEMM for each of the five specific causes of as GEMM 5-COD when referring to the models used to estimate the combined population-attributable fraction based on the five separate causes.

We refit GEMM NCD+LRI and each cause-specific GEMM without the Chinese Male Cohort (10) to examine the sensitivity
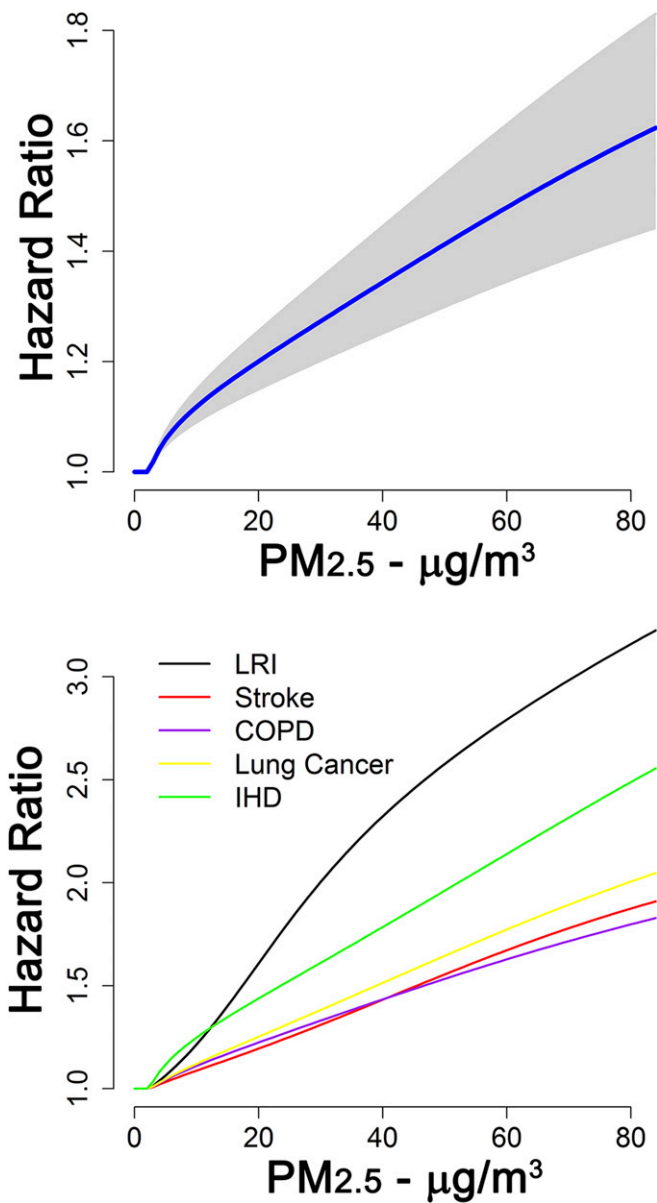
ENVIRONMENTAL SCIENCES

Fig. 1. GEMM hazard ratio predictions over $PM_{2.5}$ exposure range for noncommunicable diseases plus LRIs (NCD+LRI). (*Top*) With 95% confidence interval (gray shaded area). (*Bottom*) GEMM predictions for each of the five causes of death displayed. GEMM NCD+LRI, GEMM IHD, and GEMM stroke were based on the 60- to 64-y-old age group.

of our model predictions to this cohort, given that it incorporated a much larger range in exposure than any other study (15–84 μg/m³). We fit age-specific GEMMs for NCD+LRI, IHD, and stroke mortality (*SI Appendix*, *SI Methods*) as cardiovascular risk factors, including $PM_{2.5}$, decline with age (4).

We estimated excess mortality rates and deaths associated with a 100% reduction in 2015 $PM_{2.5}$ exposures (34) based on the GEMM and the IER model for each country and globally, a burden analysis. We also examined a partial reduction in exposure of 20%, a benefits analysis. $PM_{2.5}$ exposure estimates for 2015 were derived at a 0.1° by 0.1° grid globally based on fusion of satellite based remote sensing information, chemical transport model simulations, and spatially varying calibration to ground monitoring data using hierarchical Bayesian methods (34). The mathematical form of the GEMM and IER are described in *Methods*.

Results

GEMM NCD+LRI hazard ratio predictions increased with $PM_{2.5}$ concentration, displaying a supralinear association over lower exposures and then a near-linear association at higher concentrations (Fig. 1, *Top*). We used a counterfactual concentration of 2.4 μg/m³, the lowest observed concentration in any of the 41 cohorts (*SI Appendix*, Table S1). Below the counterfactual, we assumed no change in the hazard ratio. GEMM LRI and IHD hazard ratio predictions were larger than those for COPD, lung cancer, and stroke, which were similar to each other (Fig. 1, *Bottom*).

GEMM hazard ratio predictions were larger than those of the IER for all concentrations examined except for concentrations <10 μg/m³ for stroke with predictions declining with age for all models (*SI Appendix*, Fig. S1). In each country, the excess $PM_{2.5}$ mortality rate (deaths per 100,000 population) was calculated as the product of the cause-specific baseline mortality rate and population-attributable fraction (1 minus the inverse of the hazard ratio function) for the age-adjusted GEMM NCD+LRI, GEMM 5-COD, and IER models. For each model, we used the same estimates of exposure and baseline mortality rates. Thus, any differences were only due to the choice of hazard ratio model. The larger GEMM hazard ratio predictions resulted in higher country-specific estimates of the excess mortality rates compared with the IER-based estimates (Fig. 2). However, the correlation between excess rate estimates for the IER and GEMM NCD+LRI was high (0.95) and higher still between the IER and GEMM 5-COD (0.98).

We applied the GEMM and IER models for each country and globally to assess the excess mortality burdens related to 2015 exposure estimates (34), a 100% reduction in exposure in each country to the counterfactual exposure, and for a partial rollback of concentrations by 20% in each country, equivalent to achieving the WHO $PM_{2.5}$ air-quality first Interim Target of 35 μg/m³ at the global level (benefit analysis). We report these results grouped by global regions (Table 1). The population-weighted average $PM_{2.5}$ concentrations vary among groupings of countries from the lowest in Canada/United States (7.9 μg/m³) and Oceania (8.0 μg/m³) to the highest in China (57.5 μg/m³) and
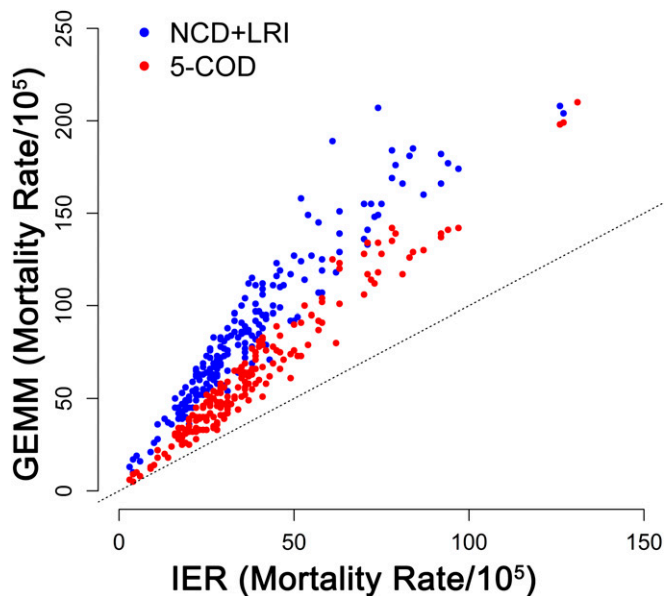

Fig. 2. Country-specific estimates of excess mortality rates associated with 100% reduction to the counterfactual concentration in population-weighted country average fine particulate matter concentrations by age-adjusted GEMM NCD+LRI vs. IER (blue dots) and GEMM 5 Causes of Death (COD) vs. IER (red dots). Dotted line represents 1:1 association.

Table 1. Population-weighted average 2015 PM$_{2.5}$ concentrations by country groupings, excess deaths (in thousands) for a 100% and 20% reduction in exposure based on GEMM NCD+LRI, GEMM 5-COD, and IER

Region	Rollback, %	PM$_{2.5}$ exposure, µg/m^3	GEMM NCD+LRI	GEMM 5-COD	IER	Ratio: IER to GEMM NCD+LRI	Ratio: GEMM 5-COD to GEMM NCD+LRI
Canada, USA	100	7.9	213	121	95	0.45	0.57
	20		42	28	20	0.48	0.68
Caribbean	100	20.2	39	28	17	0.44	0.70
	20		6	5	2	0.32	0.91
Latin America	100	17.5	365	228	152	0.42	0.63
	20		58	47	19	0.33	0.81
Africa	100	36.1	691	517	280	0.41	0.75
	20		111	102	34	0.31	0.92
Western Europe	100	13.4	439	245	176	0.40	0.56
	20		70	50	34	0.34	0.71
Eastern Europe	100	23.2	208	154	99	0.48	0.74
	20		32	28	10	0.32	0.88
Russia and EIT*	100	21.8	457	402	257	0.56	0.88
	20		70	72	26	0.37	1.03
Middle East	100	62.0	428	318	166	0.39	0.74
	20		65	56	15	0.24	0.86
China	100	57.5	2,470	1,946	1,110	0.45	0.79
	20		409	368	122	0.30	0.90
India	100	74.0	2,219	1,867	1,022	0.46	0.84
	20		359	329	108	0.30	0.92
Asia (other)	100	39.1	1,367	1,053	620	0.45	0.77
	20		216	203	69	0.32	0.94
Oceania	100	8.0	18	11	7	0.41	0.60
	20		4	3	2	0.58	0.69
Global	100	43.7	8,915	6,889	4,002	0.45	0.58
	20		1,443	1,283	452	0.31	0.89

Ratio of excess deaths between IER to GEMM NCD+LRI and GEMM 5-COD to GEMM NCD+LRI also presented.
*EIT, Economics in Transition as former Soviet states.

India (74.0 µg/m^3). Estimates of the number of excess (averted) deaths based on the GEMM NCD+LRI model were greater than the GEMM 5-COD model and still greater than the IER model for each group of countries, assuming either a 100% exposure reduction or a partial reduction of 20% (Table 1). However, the ratio of averted deaths based on the GEMM 5-COD model to the GEMM NCD+LRI model increased from between the 20% and 100% reduction scenarios, suggesting that the GEMM 5-COD model was capturing a higher percentage of the GEMM NCD+LRI averted deaths for smaller reductions in exposure. The corresponding ratios between either the IER or GEMM 5-COD to the GEMM NCD+LRI decreased from the 100% to 20% reduction scenarios in all regions except the two with the lowest exposures, suggesting that the IER-based estimates were capturing fewer of the GEMM NCD+LRI predicted averted deaths at higher concentrations.

We next examined the sensitivity of the GEMM hazard ratio predictions to the inclusion/exclusion of the Chinese cohort that covered much of the global exposure distribution. The GEMM NCD+LRI was insensitive to the exclusion of the Chinese cohort, as were the GEMM COPD and lung cancer models (*SI Appendix*, Fig. S6). However, both the IHD and stroke GEMM predictions were lower if the Chinese cohort data were not included in the model fitting (*SI Appendix*, Fig. S6). Country-specific estimates of the excess mortality rates were almost perfectly correlated between models, including and excluding the Chinese cohort (0.998), with a 14% average reduction in the estimate of the excess mortality rate among countries when the cohort was not included. Additional sensitivity analyses are presented in *SI Appendix*.

Globally, the GEMM NCD+LRI estimates 8.9 million avoided deaths [95% confidence interval (CI): 7.5–10.3] deaths in 2015.

The GEMM 5-COD estimates 6.9 million avoided deaths (95% CI: 4.9–8.5), and the IER estimates 4.0 million avoids deaths (95% CI: 3.3–4.8). We note that even in those countries from which most of the cohorts were conducted (Canada, United States, and Western Europe), the ratio of averted deaths from the IER to GEMM NCD+LRI was 0.40–0.45, based on a 100% exposure reduction (Table 1). A similar ratio was observed globally (0.45).

Discussion

To address the disease burden attributable to outdoor air pollution, governments and policymakers around the world need accurate estimates of exposure–response functions that relate changes in outdoor air pollution concentrations to changes in health risks. To date, the IER has served this purpose, but this method has several limitations, including the use of exposure/health-risk data from sources other than outdoor air pollution. Here, we demonstrated that stable hazard ratio predictions can be obtained across the global range of PM$_{2.5}$ concentrations using only studies of outdoor air pollution using an alternative hazard ratio model and method of statistical inference (25) compared to that used for the IER (3,4). By using a common statistical approach to characterizing the shape of exposure–response relationships within cohorts and then combined across cohorts, the GEMM provides a detailed examination of the shape of the outdoor exposure–mortality association, spanning the global distribution of exposure. Importantly, the manner in which we constructed the GEMM and characterized its uncertainty (*SI Appendix, SI Methods* and Fig. S7) can be directly implemented in currently available computer software used for air-quality health impact assessments, such as those used by the

US Environmental Protection Agency (USEPA) (9), Health Canada (35), and the WHO (36).

One of the most important implications of our method is that the GEMM predicted mortality hazard ratios that were almost always larger than those of the previous IER model, with much larger risks observed at higher $PM_{2.5}$ concentrations (*SI Appendix*, Fig. S1). Specifically the global estimates of mortality attributable to ambient fine particulate air pollution (8.9 million, 95% CI: 7.5–10.3) were 120% higher than previous estimates and suggest comparable impact to the leading global mortality risk factors of diet (10.3 million deaths, 95% CI: 8.8–11.9) and cigarette smoking (6.3 million deaths; 95% CI: 5.7–7.0) (1). The GEMM estimates also suggested that health benefits associated with reductions in $PM_{2.5}$ concentrations are much greater than previously suggested, particularly in areas with elevated concentrations such as India or China (Table 1).

In particular, the IER displayed the most curvature for IHD and stroke, in part due to the inclusion of hazard ratios for active smoking, which are proportionally not much larger than those for outdoor air pollution but are assigned much higher $PM_{2.5}$ exposures (4). Since the GEMM does not rely on information related to active smoking, it is not influenced by these patterns. Similarly, the GEMM does not rely on information from secondhand smoking or household heating and cooking studies. Collectively, these additional sources of exposure information included in the IER reduce hazard ratio estimates compared with the GEMM method, which relies only on data from cohort studies of outdoor air pollution.

A second important feature of our GEMM method relates to the fact that it incorporates outdoor air pollution data across the most of the global exposure range, covering 97% of the global population, owing to the inclusion of a cohort study in China (10). Importantly, our sensitivity analyses suggest that the GEMM was not sensitive to our selection of an ensemble of two models from the Chinese cohort for NCD+LRI causes, but was somewhat sensitive for IHD and stroke mortality (*SI Appendix*, Fig. S6). Moving forward, it is important that additional cohort studies be conducted in these higher-exposure environments to corroborate the results of the Chinese cohort with regard to both the shape and the magnitude of health risks associated with $PM_{2.5}$.

Our detailed analyses using subject-level data in 15 cohorts also provides direct evidence to characterize the shape of the exposure–response relationship at relatively low concentrations, an innovation of direct relevance to setting of air-quality standards. Ten of the 15 cohorts had exposures less than the WHO ambient air-quality guideline of 10 μg/m³, and in each case we observed an increase in the hazard ratio between their respective minimum concentrations and 10 μg/m³. Such evidence would not have been possible without these detailed within-cohort analyses. In comparison, the GBD2015 version of the IER model incorporated a counterfactual uncertainty distribution characterized by a uniform random variable with lower/upper bounds of 2.4 and 5.9 μg/m³, respectively. These limits were based on the average of the minimum and fifth percentiles of the exposure distributions among cohorts with relatively low concentrations (3). This counterfactual distribution was intended to describe uncertainty in the shape at low concentrations given absence of direct evidence.

Traditionally, quantitative estimates of the global mortality impacts of outdoor air pollution have been based on five specific causes of death, including lung cancer, IHD, COPD, stroke, and LRI. In this study, estimates of excess deaths for NCDs plus LRIs (NCD+LRI) were 30% higher than those based on the five specific causes of death using the GEMM. This was due, in part, to the lower baseline mortality rate for the five specific causes of death compared with NCD+LRI (52%). However, this observation also suggests that exposure to $PM_{2.5}$ is contributing to mortality from causes other than the five examined here and in the GBD (1). This is an interesting finding which supports emerging evidence that other diseases not yet included in most impact analyses are related to $PM_{2.5}$ exposure (37–39).

In summary, the GEMM method presented in this study addresses many of the limitations associated with the previous IER model and provides a means of quantifying the health impacts of outdoor air pollution. Importantly, this approach suggests that the health benefits of reducing $PM_{2.5}$ are likely much larger than previously assumed, owing to much stronger relationships between air pollution and mortality at higher concentrations. The implications of this finding are particularly significant for countries with the highest air-pollution concentrations, as the potential health benefits of air-quality improvements in these areas are larger than previously recognized.

Methods

We describe here he mathematical form of the IER and GEMM.

The IER model has the form $IER(z) = 1 + \pi(1 - \exp\{-\varphi z^\delta\})$, where $z = \max(0, PM_{2.5} - cf)$ with $cf \sim U(2.4, 5.9)$ denoting a uniform uncertainty distribution for the counterfactual, assuming no association <2.4 μg/m³. The maximum hazard ratio is $1 + \pi$, with the rate of increase for low concentrations governed by φ and for higher concentrations by δ. The unknown parameters are estimated by Bayesian methods, assuming noninformative gamma distributed priors using the computer program Stan (40). The IER was designed to estimate health burden associated not only with ambient $PM_{2.5}$ exposures, but also secondhand smoke and household air pollution; thus, the inclusion of risk information from these other particle types. IERs can take sublinear, near-linear, supralinear, and sigmodal shapes depending on the values of these parameters. For the 2015 version of the IER, a random effects error structure was assumed with random effects specific to each particle source (outdoor air pollution, secondhand smoke, household air pollution, and active smoking) (3). This additional risk information assisted in obtaining more stable risk predictions with narrower uncertainty intervals under the fully Bayesian modeling framework.

Standard computer software is not available to estimate the unknown IER parameters under a frequentist framework for survival models when examining subject-level cohort data. A Bayesian Monte Carlo approach, such as that used in Stan, is not always practical to use when the cohort is large due to computer processing limitations. We therefore needed to develop an alternative hazard ratio model and method of statistical inference.

We motivated the development of the GEMM through the Log-Linear (LL) model, as this is the most commonly used model to estimate excess deaths from exposure to ambient $PM_{2.5}$ (8, 9). The LL model has the form $LL(z) = \exp\{\beta z\}$, where $z = \max(0, PM_{2.5} - cf)$, with cf representing the counterfactual $PM_{2.5}$ concentration assuming no association below cf and unit hazard ratio when $PM_{2.5} = cf$. GEMM is an extension of the LL model by including nonlinear shapes defined by transformations, $T(z)$, of concentration. Our model has the form: $GEMM(z) = \exp\{\theta T(z)\}$. We consider transformations that cover the variety of shapes modeled by the IER, which we also suggest are useful for health impact assessment. We describe two forms of the model, one when analyzing within cohort information and another for pooling hazard ratio predictions among cohorts.

The association between concentrations of $PM_{2.5}$ and mortality for the analysis of a specific cohort is described by a class of hazard ratio functions (25): $R(z) = \exp\{\theta T(z)\}$, where $T(z) = f(z)\omega(z)$, with $f(z) = z$ or $f(z) = \log(z + 1)$, such that $R(z) = 1$ when $z = 0$ for either form of $f(z)$. Here, $\omega(z) = 1/(1 + \exp\{-(z - \mu)/(\tau r)\})$ is a logistic weighting function of z and two parameters (μ, τ) with r representing the range in the pollutant concentrations. The parameter τ controls the amount of curvature in ω with μ controlling the shape. The set of values of (f, μ, τ) define a shape of the mortality–$PM_{2.5}$ association. The estimation method is based on a routine that selects multiple values of (f, μ, τ), and, given these values, estimates of θ and its SE are obtained by using standard computer software that fit the Cox proportional hazards model (2). We can use standard computer software since we have formulated the estimation problem as a transformation of concentration, $T(z) = f(z)\omega(z)$, and a single unknown parameter θ. An ensemble model is calculated by the weighted average of the predictions of all models examined at any concentration with weights defined by the likelihood function value. Uncertainty estimates of the ensemble model predictions are obtained by bootstrap methods, which incorporate both sampling and model shape uncertainty (25). For large negative values of μ, $\omega(z) \sim 1$, and in such cases, $T(z) \sim z$ when $f(z) = z$, and $T(z) \sim \log(z + 1)$ when $f(z) = \log(z + 1)$. We thus obtain a family of shapes including approximately linear, log linear, supralinear and sublinear, and S-shaped in concentration. Details of the set of values of (f, μ, τ) and the estimation routine are described elsewhere (25).

We define a modification of the hazard ratio model used for the analysis of subject level data within each cohort as our common model among cohorts: $R(z) = \exp\{\theta T(z)\}$, where $T(z) = \log(1 + z/\alpha)\omega(z)$. We have replaced the two forms of $f(z)$ that we used in the analysis of the subject level within cohort data by a single mathematical form $\log(1 + z/\alpha)$ defined by an additional parameter α. Here, α controls the amount of curvature in R with less curvature for larger values of α. For larger values of α, the model is near linear for low concentrations and for low values of μ. However, changes in the hazard ratio decline with increasing concentrations beyond the range of the data (*SI Appendix*, Fig. S9). We do this so that predictions of the hazard ratio beyond the observed exposure range have a logarithmic form with diminishing changes in association as exposure increases. This structure limits the size of the predicted hazard ratio over concentration ranges where we have no observations. Estimates of $R(z)$ are obtained by specifying values in the parameters (α, μ, τ) that define the shape of the transformations and given these shapes, θ is estimated by using standard computer software. An ensemble estimate is then constructed of all of the shapes examined weighted by their respective likelihood values. Bootstrap methods were used to obtain uncertainty intervals (*SI Appendix, SI Methods*). We constrained the amount of curvature in our fitted model by restricting the selection of values of (α, μ, τ) (*SI Appendix, SI Methods*).

The ensemble model predictions are a weighted average of all models examined. To simplify the presentation of our results and their use for burden/benefits analyses, we suggest fitting a single algebraic function of the same form as the GEMM to the ensemble model predictions over the concentration range of interest for health burden assessment. In this work, we use the concentration interval 2.4–84 μg/m³ by 0.1-μg/m³ increments. We simplified the model somewhat by absorbing the concentration range r with the parameter τ by setting $r\tau = \nu$. We then estimated the parameters by standard nonlinear regression methods [R routine nlsLM from the package "minpack.lm" (41)]. We attribute all of the uncertainty in the ensemble model prediction to θ by first regressing the SE of the $\log R(z)$ predictions among the bootstrap samples at each concentration on the $\log R(z)$ predictions themselves. The slope of this regression is designated as the SE of our estimate of θ. The adequacy of our model approximation is examined by plotting the ensemble and uncertainty intervals overlaid with the approximate model average and approximate uncertainty intervals (*SI Appendix*, Fig. S7).

Since the hazard ratio declines with age for cardiovascular risk factors, such as PM₂.₅, we constructed age-specific GEMMs in a manner similar to that used by the GBD program (1) (*SI Appendix, SI Methods*). We did this for the GEMM NCD+LRI and the IHD and stroke GEMMs. We only age-adjusted the proportion of cardiovascular deaths in each of the 41 cohorts for age for the GEMM NCD+LRI (*SI Appendix, SI Methods*). Thus, age-adjusted GEMM NCD+LRI displayed less variation than those for IHD and stroke (*SI Appendix*, Fig. S1).

Estimates of excess deaths were determined by the product of the size of country population, the age-specific annual mortality rate, and the population attribution fraction (one minus the inverse of the relative risk function). We interpret the hazard ratio functions obtained from cohort studies as relative risks to use the population attributable risk definition since the annual baseline mortality rates are small (*SI Appendix*).

1. GBD 2015 Risk Factors Collaborators (2016) Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: A systematic analysis for the global burden of disease study 2015. *Lancet* 388:1659–1724.
2. Cox DR (1972) Regression models and life-tables. *J R Stat Soc B* 34:187–220.
3. Cohen A, et al. (2017) Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the global burden of diseases study 2015. *Lancet* 389:1907–1918.
4. Burnett RT, et al. (2014) An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environ Health Perspect* 122:397–403.
5. World Health Organization (2014) Mortality and burden of disease from ambient air pollution: Situation and trends. Available at www.who.int/gho/phe/outdoor_air_pollution/burden_text/en/. Accessed January 14, 2018.
6. Qin Y, et al. (2017) Air quality, health, and climate implications of China's synthetic natural gas development. *Proc Natl Acad Sci USA* 114:4887–4892.
7. Lacey FG, Henze DK, Lee CJ, van Donkelaar A, Martin RV (2017) Transient climate and ambient health impacts due to national solid fuel cookstove emissions. *Proc Natl Acad Sci USA* 114:1269–1274.
8. US Environmental Protection Agency (2012) Regulatory impact analysis for the final revisions to the national ambient air quality standards for particulate matter (Office of Air Quality Planning and Standards, Health and Environmental Impacts Division, Research Triangle Park, NC), Technical Report EPA-452/R-12-005.
9. US Environmental Protection Agency (2015) Environmental Benefits Mapping Analysis Program Community Edition (BenMAP-CE). Available at https://www.epa.gov/benmap. Accessed November 16, 2017.
10. Yin P, et al. (2018) Long-term exposure to fine particulate matter and cardiovascular disease mortality in China: A cohort study. *Environ Health Perspect*, 10.1289/EHP1673.
11. Turner MC, et al. (2016) Long-term ozone exposure and mortality in a large prospective study. *Am J Respir Crit Care Med* 193:1134–1142.
12. Thurston GD, et al. (2016) Ambient particulate matter air pollution exposure and mortality in the NIH-AARP Diet and Health cohort. *Environ Health Perspect* 124:484–490.
13. Carey IM, et al. (2013) Mortality associations with long-term exposure to outdoor air pollution in a national English cohort. *Am J Respir Crit Care Med* 187:1226–1233.
14. Villeneuve PJ, et al. (2015) Long-term exposure to fine particulate matter air pollution and mortality among Canadian women. *Epidemiology* 26:536–545.
15. Hart JE, et al. (2015) The association of long-term exposure to PM2.5 on all-cause mortality in the Nurses' health study and the impact of measurement-error correction. *Environ Health* 14:38–36.
16. Lipsett MJ, et al. (2011) Long-term exposure to air pollution and cardiorespiratory disease in the California teachers study cohort. *Am J Respir Crit Care Med* 184:828–835.
17. Pinault LL, et al. (2017) Associations between fine particulate matter and mortality in the 2001 Canadian census health and environment cohort. *Environ Res* 159:406–415.
18. Crouse DL, et al. (2015) Associations between ambient PM₂.₅, O₃, and NO₂ and mortality in the Canadian census health and environment cohort (CanCHEC) over a 16-year follow-up. *Environ Health Perspect* 123:1180–1186.
19. Pinault L, et al. (2016) Risk estimates of mortality attributed to low concentrations of ambient fine particulate matter in the Canadian community health survey cohort. *Environ Health* 15:18–31.
20. Cesaroni G, et al., (2013) Long-term exposure to urban air pollution and mortality in a cohort of more than a million adults in Rome. *Environ Health Perspect* 121:324–331.
21. Wong CM, et al. (2015) Satellite-based estimates of long-term exposure to fine particles and association with mortality in elderly Hong Kong residents. *Environ Health Perspect* 123:1167–1172.
22. Pope CA, III, et al. (2017) Mortality risk and PM₂.₅ air pollution in the United States: An analysis of a national prospective cohort. *Air Qual Atmos Health* 11:245–252.
23. Fischer PH, et al. (2015) Air pollution and mortality in seven million adults: The Dutch environmental longitudinal study (DUELS). *Environ Health Perspect* 123:697–704.
24. Beelen R, et al. (2014) Effects of long-term exposure to air pollution on natural-cause mortality: An analysis of 22 European cohorts within the multicentre ESCAPE project. *Lancet* 383:785–795.
25. Nasari M, et al. (2016) A class of non-linear exposure-response models suitable for health impact assessment applicable to large cohort studies of ambient air pollution. *Air Qual Atmos Health* 9:961–972.
26. Lepeule J, Laden F, Dockery D, Schwartz J (2012) Chronic exposure to fine particles and mortality: An extended follow-up of the Harvard six cities study from 1974 to 2009. *Environ Health Perspect* 120:965–970.
27. Puett RC, Hart JE, Suh H, Mittleman M, Laden F (2011) Particulate matter exposures, mortality, and cardiovascular disease in the health professionals follow-up study. *Environ Health Perspect* 119:1130–1135.
28. Weichenthal S, et al. (2014) Long-term exposure to fine particulate matter: Association with nonaccidental and cardiovascular mortality in the agricultural health study cohort. *Environ Health Perspect* 122:609–615.
29. Beelen R, et al. (2008) Long-term effects of traffic-related air pollution on mortality in a Dutch cohort (NLCS-AIR study). *Environ Health Perspect* 116:196–202.
30. McDonnell WF, Nishino-Ishikawa N, Petersen FF, Chen LH, Abbey DE (2000) Relationships of mortality with the fine and coarse fractions of long-term ambient PM10 concentrations in nonsmokers. *J Expo Anal Environ Epidemiol* 10:427–436.
31. Tseng E, et al. (2015) Chronic exposure to particulate matter and risk of cardiovascular mortality: Cohort study from Taiwan. *BMC Public Health* 15:936–945.
32. Di Q, et al. (2017) Air pollution and mortality in the Medicare population. *N Engl J Med* 376:2513–2522.
33. Bentayeb M, et al. (2015) Association between long-term exposure to air pollution and mortality in France: A 25-year follow-up study. *Environ Int* 85:5–14.
34. Shaddick G, et al. (2018) Data integration model for air quality: A hierarchical approach to the global estimation of exposures to ambient air pollution. *J R Stat Soc Ser C* 67:231–253.
35. Judek S, Stieb D, Jovic B, Edwards B (2012) Air quality benefits assessment tool user guide (Health Canada, Ottawa). Available at science.gc.ca/eic/site/063.nsf/eng/h_97170.html. Accessed December 7, 2017.
36. World Health Organization (2018) AirQ+: Software tool for health risk assessment of air pollution. Available at www.euro.who.int/en/health-topics/environment-and-health/air-quality/activities/airq-software-tool-for-health-risk-assessment-of-air-pollution. Accessed January 12, 2018.
37. Eze IC, et al. (2015) Association between ambient air pollution and diabetes mellitus in Europe and North America: Systematic review and meta-analysis. *Environ Health Perspect* 123:381–389.
38. Thurston GD, et al. (2017) A joint ERS/ATS policy statement: What constitutes an adverse health effect of air pollution? An analytical framework. *Eur Respir J* 49:1600419.
39. Chen H, et al. (2017) Exposure to ambient air pollution and the incidence of dementia: A population-based cohort study. *Environ Int* 108:271–277.
40. Carpenter B, et al. (2017) Stan: A probabilistic programming language. *J Stat Softw*, 10.18627/jss.v076.i01.
41. R Core Team (2016) R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna). Available at https://www.R-project.org/.

ENVIRONMENTAL SCIENCES