

## Supplementary Information

### **Global Estimates of Mortality Associated with Long Term Exposure to Outdoor Fine Particulate Matter**

Richard Burnett<sup>1</sup>, Hong Chen<sup>1,2</sup>, Mięczysław Szyszkwicz<sup>1</sup>, Neal Fann<sup>3</sup>, Bryan Hubbell<sup>4</sup>, C. Arden Pope III<sup>5</sup>, Joshua S. Apte<sup>6</sup>, Michael Brauer<sup>7</sup>, Aaron Cohen<sup>8</sup>, Scott Weichenthal<sup>9</sup>, Jay Coggins<sup>10</sup>, Qian Di<sup>11</sup>, Bert Brunekreef<sup>12</sup>, Joey Frostad<sup>13</sup>, Stephen S. Lim<sup>13</sup>, Haidong Kan<sup>14</sup>, Katherine D. Walker<sup>8</sup>, George D. Thurston<sup>15</sup>, Richard B. Hayes<sup>16</sup>, Chris C. Lim<sup>17</sup>, Michelle C. Turner<sup>18</sup>, Michael Jerrett<sup>19</sup>, Daniel Krewski<sup>20</sup>, Susan Gapstur<sup>21</sup>, W. Diver<sup>21</sup>, Bart Ostro<sup>22</sup>, Debbie Goldberg<sup>23</sup>, Daniel L. Crouse<sup>24</sup>, Randall Martin<sup>25</sup>, Paul Peters<sup>26</sup>, Lauren Pinault<sup>27</sup>, Michael Tjepkema<sup>27</sup>, Aaron van Donkelaar<sup>25</sup>, Paul J. Villeneuve<sup>28</sup>, Anthony B. Miller<sup>29</sup>, Peng Yin<sup>30</sup>, Maigeng Zhou<sup>30</sup>, Lijun Wang<sup>30</sup>, Nicole A.H. Janssen<sup>31</sup>, Marten Marra<sup>31</sup>, Richard W. Atkinson<sup>32</sup>, Hilda Tsang<sup>33</sup>, Thuan Quoc Thach<sup>33</sup>, John Cannon<sup>5</sup>, Ryan Allen<sup>5</sup>, Jaime E. Hart<sup>34</sup>, Francine Laden<sup>34</sup>, Giulia Cesaroni<sup>35</sup>, Francesco Forastiere<sup>35</sup>, Gudrun Weinmayr<sup>36</sup>, Andrea Jaensch<sup>36</sup>, Gabriele Nagel<sup>36</sup>, Hans Concin<sup>37</sup>, Joseph V. Spadaro<sup>38</sup>

<sup>1</sup> Population Studies Division, Health Canada, Ottawa, Canada

<sup>2</sup> Public Health Ontario, Toronto, Ontario M5G 1V2

<sup>3</sup> USEPA, Office of Air Quality Planning & Standards, Risk and Benefits Group

<sup>4</sup> USEPA Office of Research and Development

<sup>5</sup> Department of Economics, Brigham Young University, Provo, UT, U.S

<sup>6</sup> Department of Civil, Architectural and Environmental Engineering. University of Texas at Austin.

<sup>7</sup> School of Population and Public Health, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada

<sup>8</sup> Health Effects Institute, 75 Federal Street Suite 1400, Boston, MA 02110-1817, USA

<sup>9</sup> Department of Epidemiology, Biostatistics, and Occupational Health and Gerald Bronfman Department of Oncology, McGill University

<sup>10</sup> Department of Applied Economics. University of Minnesota

<sup>11</sup> Department of Biostatistics. Harvard T.H. Chan School of Public Health

<sup>12</sup> Institute for Risk Assessment Sciences, Universiteit Utrecht

<sup>13</sup> Institute for Health Metrics and Evaluation, University of Washington

<sup>14</sup> School of Public Health, Fudan University, Shanghai, China.

<sup>15</sup> Environmental Medicine and Population Health, Program in Human Exposures and Health Effects, New York University School of Medicine

<sup>16</sup> Department of Population Health, NYU Langone Medical Center

<sup>17</sup> New York University School of Medicine

<sup>18</sup> ISGlobal, Barcelona Institute for Global Health, Barcelona Biomedical Research Park (PRBB)

<sup>19</sup> Department of Environmental Health Sciences, Fielding School of Public Health, University of California, Los Angeles

<sup>20</sup> McLaughlin Centre for Population Health Risk Assessment, University of Ottawa

<sup>21</sup> American Cancer Society, Inc.

<sup>22</sup> University of California, Davis

<sup>23</sup> Cancer Prevention Institute of California

<sup>24</sup> Department of Sociology, University of New Brunswick

<sup>25</sup> Department of Physics and Atmospheric Science, Dalhousie University, Halifax

<sup>26</sup> Departments of Sociology and Economics and Canada Research Chair, Spatial and Social Inequalities of Health, University of New Brunswick

<sup>27</sup> Health Analysis Division, Statistics Canada

<sup>28</sup> Department of Health Sciences, Carleton University, Ottawa Canada K1S 5B6

<sup>29</sup> Dalla Lama School of Public Health, University of Toronto, Toronto, Canada

<sup>30</sup> National Center for Chronic Noncommunicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China

<sup>31</sup> National Institute for Public Health and the Environment, PO Box 1, 3720 BA Bilthoven, The Netherlands

<sup>32</sup> Population Health Research Institute and MRC-PHE Centre for Environment and Health, St George's, University of London

<sup>33</sup> School of Public Health, University of Hong Kong, China

<sup>34</sup> Department of Environmental Health, Harvard C.T. Channing School of Public Health, Harvard, Boston

<sup>35</sup> Department of Epidemiology of the Regional Health Service, ASL Roma 1, Italy.

<sup>36</sup> Institute of Epidemiology and Medical Biometry, Ulm University

<sup>37</sup> Agency for Preventive and Social Medicine (aks), Rheinstraße 61, 6900 Bregenz, Austria

<sup>38</sup> Private Consultant, Bilbao, Spain

## SI Methods

Here, we describe the statistical estimation of the unknown parameters in our new hazard ratio function. Two forms of the model are examined: the first is used for the analyses of the 15 selected cohorts on primary subject-level data; the second is more general and is used to identify a common hazard ratio model among all 41 cohorts. The first more simplified model is used to reduce the computational burden in the analysis of primary cohort data.

**Hazard ratio model estimation for cohorts with subject-level data.** The association between concentrations of air pollution,  $z$ , and mortality for the analysis of a specific cohort is described by a class of hazard ratio functions (25):  $R(z)=\exp\{\theta f(z)\omega(z)\}$ , where  $f(z)=z$  or  $f(z)=\log(z+1)$ , such that  $R(z)=1$  when  $z=0$  for either form of  $f$ . Here,  $\omega(z)=1/(1+\exp\{-(z-\mu)/(\tau r)\})$  is a logistic weighting function of  $z$  and two parameters  $(\mu, \tau)$  with  $r$  the range in the pollutant concentrations. The parameter  $\tau$  controls the amount of curvature in  $\omega$  with  $\mu$  controlling the shape. The set of values of  $(f, \mu, \tau)$  define a shape of the mortality-PM<sub>2.5</sub> association. The estimation method is based on a routine that selects multiple values of  $(f, \mu, \tau)$  and given these values, estimates of  $\theta$  and its standard error are obtained 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,  $f(z)\omega(z)$ , and a single unknown parameter  $\theta$ . 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).

We applied this modeling procedure to the 15 cohorts for which we have access to the individual subject-level data (10-24). These cohorts have a large number of deaths resulting in increased statistical power to detect differences in the shapes of the PM<sub>2.5</sub>-mortality associations. In order to represent the ensemble curve for each of these 15 cohorts, we first trim the exposure data by removing values below the 1<sup>st</sup> and above the 99<sup>th</sup> percentiles as model uncertainty can be substantially larger when these extreme exposure values are included. We then calculate the ensemble model prediction at 101 equally spaced concentrations of the trimmed exposure distribution after shifting these exposures such that the minimum value is zero (i.e. subtracting the minimum concentration from all exposures). The resulting predicted values represent the hazard ratio at each of 100 values compared to the minimum concentration, whose hazard ratio is unity, for each study separately. In addition we calculate the statistical uncertainty associated with each prediction.

For the 15 cohorts where we examined the shapes of concentration-mortality associations in detail, we also summarized error by a single measure. Each of the 100 predictions is perfectly correlated given the ensemble model. We first calculated a 100 by 100 covariance matrix assuming perfect correlation and summed the rows of the matrix. We then determined the maximum of these sums as our single measure of uncertainty and assigned it to each of the 100 predictions thus forming a diagonal covariance matrix for each study. This approach avoids the issue where predicted values close to the minimum concentration have smaller uncertainty by definition and thus does not assign undue weight to these values.

For the remaining 26 cohorts (24-33) we extracted the logarithm of the hazard ratio from the published literature along with associated standard errors. Following the GBD 2015 (1) approach, we calculated the hazard ratio and standard error for each cohort for an exposure contrast between the 5<sup>th</sup> and 95<sup>th</sup> percentiles. We thus extracted a single measure of association and error from each study.

**Global Exposure Mortality Model.** We propose a multivariate random effects model to pool the hazard ratio functions among the 41 cohorts. The data for the model is denoted by  $r_{si}$  which is the predicted logarithm of the hazard ratio (log-HR) for the  $i^{\text{th}}$  exposure contrast in the  $s^{\text{th}}$  study with corresponding concentration  $z_{si}$ . Fifteen of the 41 cohorts have 100  $r_{si}$  values and the remaining 26 cohorts have a single value. We also define  $z_{s0}$  for the 15 studies such that  $R(z_0)=1$ . For the remaining 26 cohorts,  $z_{s0}$  is defined as the 5<sup>th</sup> percentile and  $z_{s1}$  is the 95<sup>th</sup> percentile.

Consider a common relative risk model for all cohorts of the form:  $R(z)=\exp\{\theta\log(1+z/\alpha)\omega(z)\}$ . 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  $\alpha$ . Here,  $\alpha$  controls the amount of curvature in  $R$  with less curvature for larger values of  $\alpha$ . We do this so that predictions of relative risk beyond the observed exposure range have a logarithmic form with diminishing changes in relative risk as exposure increases. This structure limits the size of the predicted relative risks over concentration ranges where we have no observations (SI Appendix, Figure S9).

The standard error of the predicted log-HR for the 15 cohorts to which we applied the non-linear hazard ratio models is a function of both sampling error and model uncertainty. However, in the 26 cohorts where we did not have access to the primary data and thus assumed a linear in concentration hazard ratio model, the uncertainty is solely a function of sampling error. In order to define similar error structures for both types of information we calculated the mean ratio of total error to sampling error in the 15 primary data cohorts and multiplied this value (1.64) by the sampling error of each of the 26 non-primary data cohorts.

We related our model  $\log(R(z_{si}))- \log(R(z_{s0}))$  to the observations  $r_{si}$  using the R routine “rma.mv” with options: method=“REML” and struct=“CS”, where CS denotes compound symmetry (41). Here, we assume a common variance and correlation in predicted log-hazard ratios between different studies. We do this by first creating a class of transformations of concentration in a similar manner to what we did for the analysis of the primary data cohorts by setting  $\alpha=(1,3,5,7,9)$ ,  $\tau=(0.1, 0.2, 0.3, 0.4, 0.5, 0.6)$ , and values of  $\mu$  representing the 0<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, and 95<sup>th</sup> percentiles of the PM<sub>2.5</sub> exposure distribution among all cohorts. These parameter specifications define 150 different shapes, from supra-linear, near linear, to sub-linear. We have found that values of  $\alpha>9$  do not appreciably alter the shape of the model nor do values of  $\tau>0.6$ . We do not want to include models that have very similar shapes as we would be giving these shapes more weight in the ensemble since they would be represented several times. For each set of the parameter values the rma.mv routine is used to calculate an estimate of  $\theta$  and its standard error. The standard error is a function of both within cohort estimation error and between cohort heterogeneity. We then calculate the ensemble estimate at any PM<sub>2.5</sub> concentration based on a weighted average of all 150 model predictions examined, with bootstrap errors (25). We also considered two additional models representing the correlation structure between cohorts: Heteroscedastic Compound Symmetry - common correlation between studies but allow the variance to vary by study; and Unstructured – both variance and correlation vary by study. GEMM parameter estimates were similar for all three between cohort error structures examined (SI Appendix, Table S3).

**Age Adjustment for Cardiovascular Mortality.** For the cardiovascular outcomes, Ischemic Heart Disease (IHD) and Stroke, GBD generates age-specific relative risk functions by adjusting each cohort-specific logarithm of the

hazard ratio for the median age of follow-up of the cohort (4). The median age, *medage*, of follow-up is estimated by the average age at commencement of follow-up plus half the length of follow-up in years. The adjustment formula is:  $(age-100)/(medage-110)$ , for any specific age. Here, when age is 110 years, no association is assumed. We apply a similar age adjustment to the hazard ratio for each cohort by using the competing risk hazard model specification:  $\lambda_N R_N(z) C_N = \lambda_{CV} R_{CV}(z) C_{CV} + \lambda_{nonCV} R_{nonCV}(z) C_{nonCV}$ , where  $\lambda_N$  is the baseline mortality rate for non-accidental causes of death (COD), with  $\lambda_{CV}$  and  $\lambda_{nonCV}$  denoting the mortality rates from cardiovascular and non-cardiovascular COD respectively. Here,  $C_N$ ,  $C_{CV}$  and  $C_{nonCV}$  are the hazard ratio functions from the other covariates in the survival model, such as smoking, for non-accidental, cardiovascular, and non-cardiovascular COD respectively. Further assuming the cardio-vascular and non-cardio-vascular models are identical to the non-accidental COD model, and the hazard ratios for the other covariates are also identical, we can write the non-accidental COD model as a weighted combination of the proportional mortality rates from cardiovascular and non-cardiovascular COD as:  $R_N(z) = p_{CV} R_{CV}(z) + (1-p_{CV}) R_{nonCV}(z)$ , with  $p_{CV} = \lambda_{CV} / \lambda_N$ . Noting that the observed hazard ratio model can be written as  $R_N(z) = \exp\{\theta T(z)\}$ , the age-adjusted risk for each cohort is:  $p_{CV} \exp\{\theta T(z)(age-110)/(medage-110)\} + (1-p_{CV}) \exp\{\theta T(z)\}$ .

For each of the 41 cohorts we determined the median age of follow-up and extracted the mortality rate from the GBD database (43)) for the time period of follow-up, proportion of males and females, and the country the cohort was conducted in for both non-accidental and cardiovascular mortality by 12 five-year age groups (i.e. 25-29, 30-34, ..., over 80 years) corresponding to those used by the GBD (43). We then fit age-specific pooled hazard ratio models to the age-adjusted hazard ratios and their standard errors using the GEMM with the same values of  $(\alpha, \mu, \nu)$  as the non-age adjusted GEMM. Thus only  $\theta$  is estimated from the age-adjusted hazard ratios. We also constructed age-specific GEMMs for IHD and Stroke mortality in a manner similar to that of GBD (1).

## SI Results

A list of the 41 cohorts, their respective areas (Canada, United States, Europe, Asia), hazard ratios per  $10\mu\text{g}/\text{m}^3$  change in  $\text{PM}_{2.5}$  concentrations assuming a log-linear model, and the exposure limits are presented in SI Appendix, Table S1. We observed hazards ratios greater than unity for 35 of the 41 cohorts. Predicted hazard ratios for each of the 15 cohorts with subject-level data analyses are presented in SI Appendix, Figure S2. The cohort-specific predictions are defined with respect to their counterfactual concentration defined by their lowest observed exposure. The vast majority (12 of 15) of hazard ratio models displayed a supra-linear association between mortality and  $\text{PM}_{2.5}$  exposure with no evidence of a sub-linear association over the lowest exposures relative to each study. However, three cohorts (Canadian National Breast Screening Cohort, Hong Kong Cohort, and the Chinese Male Cohort) did display a sub-linear association at their respective lowest exposures. The GEMM NCD+LRI is displayed in Figure 1 (upper left hand panel) based on the pooling of hazard ratio predictions in the 15 cohorts in addition to the 26 cohorts where a log-linear specification was assumed. Note that information to estimate the GEMM from each cohort is only used over the exposure range of that cohort. The pooling approach uses information both on the shape of the association within each cohort and the manner in which these shapes vary between cohorts with different exposure ranges (Methods and SI Appendix, Methods). The GEMM displays a small amount of supra-linear curvature over the lower concentrations with less curvature at higher concentrations. We assumed that there was no change in the hazard ratio for concentrations below the lowest observed exposure in any cohort ( $2.4\mu\text{g}/\text{m}^3$ ). We thus define the GEMM with respect to this value as our single counterfactual concentration.

GEMM parameter estimates for NCD+LRI and the five specific causes of death, including and excluding the Chinese cohort are given in SI Appendix, Table S2. Both the shape and magnitude of the GEMM was robust to

the manner in which we included information on the hazard ratio within the 15 cohorts where we conducted detailed analyses, and to the number of cohorts included in the pooling (SI Appendix, Figure S3). For example, GEMM predictions assuming a log-linear within cohort model using only the 15 cohorts with detailed analyses yielding similar predictions compared to also including the 26 additional cohorts (SI Appendix, Figure S3 upper left hand panel). A similar pattern of association was also observed if we assumed a log-linear model for these 15 cohorts (upper right hand panel), or assuming a log-linear model for all 41 cohorts (lower left hand panel). We did observe a slight difference if we examined only the 26 cohorts under a log-linear within cohort model with some additional uncertainty (lower right hand panel). We conclude from these analyses that a consistent association is observed between ambient fine particulate exposure and non-accidental mortality among the global cohort studies.

To estimate the excess number of deaths associated with  $PM_{2.5}$  exposure one requires estimates of exposure, the size of the population exposed, the age-specific mortality rate for that population, and the fraction of total deaths attributable to that exposure (Population Attributable Fraction - PAF). The PAF is defined as one minus the inverse of the relative risk function. The relative risk function is the ratio of the probability of death by a certain age given a specific exposure to the probability of death at that age assuming the counterfactual exposure (42). However, the information used to develop the GEMM and IER are based on hazard ratio functions that model the probability of death over a fixed time interval given that subjects were alive at the beginning of the time interval. Relative risks and hazard ratios are similar in magnitude when the mortality probability is small (42) and in such cases one can substitute the hazard ratio for the relative risk when calculating excess deaths. However, we use age specific mortality rates which are the probability of death over a short time scale, given people were alive at the start of the year. In our burden estimates we use a single year, 2015. These rates are then acting like a hazard ratio. Furthermore, they are generally small ( $<0.1$ ) for most age groups (SI Appendix, Figure S4). Here, as does the GBD (1), we make such a substitution.

Several estimates of hazard ratios were reported for the Chinese Male Cohort (10) including those based on survival models containing only covariates measured at the individual level, such as smoking, obesity and diet. A series of models were also presented including contextual covariates: urban/rural designation, major region of China, and a single covariate reporting the percentage of low education in the subject's area of residence. Detailed examination of the shape of the  $PM_{2.5}$ -mortality association was conducted for two model specifications: 1) all individual level covariates, and 2) individual plus contextual covariates (SI Appendix, Figure S5 – left hand panel). Hazard ratio predictions were similar for lower concentrations among the two model specifications but very different for higher concentrations. For our main analyses we constructed an ensemble estimate (25) of the two model predictions assuming they were equally valid. The ensemble model predictions are an average of the two specific model predictions (blue line of SI Appendix, Figure S5, left hand panel) with its uncertainty a function of the uncertainty in each model prediction and the variation between model predictions (SI Appendix, Figure S5 – left hand panel - grey shaded area). This resulted in a much wider uncertainty interval for the ensemble model compared to either of the two specific models. The influence of these three model specifications for the Chinese Male Cohort (10) on the resulting GEMM NCD+LRI model predictions is displayed in SI Appendix, Figure S5 (right hand panel). In addition, we fit a GEMM excluding the Chinese Male Cohort. The GEMM predictions were clearly larger at higher concentrations ( $> 30\mu g/m^3$ ) when using the Chinese Male Cohort with both individual and contextual covariates (SI Appendix, Figure S5, right-panel, red line). However, the GEMM predictions were almost the same for how we incorporated information from the China cohort for the individual covariate, ensemble of the two models, or excluding the China cohort. We recommend that the

most appropriate GEMM NCD+LRI model is the one incorporating the ensemble of the two models for the China cohort.

We also compared GEMMs for each of the five specific causes of death (SI Appendix, Figure S6) in addition to the NCD+LRI model to the sensitivity of exclusion of the Chinese cohort. Exclusion of the cohort had little influence for NCD+LRI, COPD, and Lung Cancer. However, the GEMM excluding the Chinese cohort yielded lower hazard ratio predictions for Stroke and IHD.

There are several computer programs that estimate public health burden from exposure to ambient air pollution, including that of the United States Environmental Protection Agency (9), Health Canada (35), and World Health Organization (36). These programs principally use the log-linear hazard ratio model formulation characterized by a single parameter estimate and uncertainty. However, the GEMM is defined by several parameters and ensemble model based uncertainty that does not directly conform to these software applications. We developed an approximation to the GEMM that allows such direct implementation (Methods) with a comparison of the ensemble model based predictions and uncertainty to our approximation (SI Appendix, Figure S7). In this case the approximation is reasonable with the mean predictions indistinguishable from each other and similar estimates of uncertainty. We used this approximation on all further estimates of excess mortality rates of deaths associated with PM<sub>2.5</sub> ambient exposures.

The number of global deaths in 2015 rapidly increased with age (SI Appendix, Figure S8 – upper panel) with the largest numbers due to IHD of the five specific causes examined. For example, for the 50-54 year old age group, 46% of NCD+LRI deaths were due to the five causes, with 19% due to IHD. However, these values increased for the over 80 year group, with 59% of NCD+LRI deaths due to the five causes and 25% due to IHD. GEMM based estimates of excess deaths due to PM<sub>2.5</sub> exposure also increase with age by cause of death in conjunction with their corresponding baseline mortality estimates (SI Appendix, Figure S8 – lower panel). Even at the older age groups there are substantial numbers of excess deaths from PM<sub>2.5</sub> exposure for causes not due to the specific five examined here.

## References

1. GBD 2015 Risk Factors Collaborators. 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. *The Lancet* 2016; 388: 1659–1724.
2. Cox, DR (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society, Series B* 34 (2): 187–220.
3. Cohen A et al., 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. Published **Online** April 10, 2017 [http://dx.doi.org/10.1016/S0140-6736\(17\)30505-6](http://dx.doi.org/10.1016/S0140-6736(17)30505-6).
4. Burnett RT et al., An Integrated Risk Function for Estimating the Global Burden of Disease Attributable to Ambient Fine Particulate Matter Exposure. *Environ Health Perspect* 2014; published online Feb 11. DOI:10.1289/ehp.1307049.
5. World Health Organization. [http://www.who.int/gho/phe/outdoor\\_air\\_pollution/burden\\_text/en/](http://www.who.int/gho/phe/outdoor_air_pollution/burden_text/en/).
6. Qin Y et al., Air quality, health, and climate implications of China's synthetic natural gas development. *Proc Natl Acad Sci U S A*. 2017 May 9;114(19):4887-4892. doi: 10.1073/pnas.1703167114.

7. Lacey FG, Henze DK, Lee CJ, van Donkelaar A, Martin RV. Transient climate and ambient health impacts due to national solid fuel cookstove emissions. *Proc Natl Acad Sci U S A*. 2017 Feb 7;114(6):1269-1274. doi: 10.1073/pnas.1612430114.
8. USEPA (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. EPA-452/R-12-005 December.
9. US EPA (2015) Environmental Benefits Mapping Analysis Program Community Edition (BenMAP-CE). <http://www2.epa.gov/benmap>
10. Yin P et al. (2018), Long-term exposure to fine particulate matter and cardiovascular disease mortality in China: a cohort study. *Environmental Health Perspectives*. <https://doi.org/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, 10, 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, 11, 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 PM<sub>2.5</sub> on all-cause mortality in the Nurses' Health Study and the impact of measurement-error correction. *Environmental 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. *American Journal of Respiratory & Critical Care Medicine*. doi:10.1164/rccm.201012-2082OC
17. Pinault LL et al., (2017). Associations between fine particulate matter and mortality in the 2001 Canadian Census Health and Environment Cohort (2001 CanCHEC). *Environmental Research*, 159: 406–415.
18. Crouse DL et al., (2015). Associations between Ambient PM<sub>2.5</sub>, O<sub>3</sub>, and NO<sub>2</sub> and Mortality in the Canadian Census Health and Environment Cohort (CanCHEC) over a 16-year Follow-Up. *Environmental Health Perspectives* 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. *Environmental 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 III CA et al., (2017). Mortality Risk and PM<sub>2.5</sub> Air Pollution in the United States: An Analysis of a National Prospective Cohort. *Air Quality and Atmospheric Health*. <https://doi.org/10.1007/s11869-017-0535-3>.
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: 785–95.
25. Nasari M et al., (2015). A Class of Non-Linear Exposure-Response Models Suitable for Health Impact Assessment Applicable to Large Cohort Studies of Ambient Air Pollution. *Air Quality, Atmosphere, and Health*: DOI: 10.1007/s11869-016-0398-z.
26. Lepeule J, Ladan 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. *Environmental Health Perspectives* 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. *Environmental Health Perspectives* 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). *Environmental Health Perspectives* 116:196–202.
30. McDonnell WF, Nishimo-Ishikawa N, Peterson FF, Chen LH, Abbey DE (2000). Relationships of Mortality with Fine and Coarse Fractions of Long-Term Ambient PM<sub>10</sub> in Non Smokers. *Journal of Exposure Analysis and Environmental Epidemiology* 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* 376:2513-22.
33. Bentayeb M et al., (2015). Association between long-term exposure to air pollution and mortality in France: A 25-year follow-up study. *Environment International* 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. *Journal of the Royal Statistical Society, Series C*, 67:231-253.
35. Judek S, Stieb D, Jovic B, Edwards B (2012) Air quality benefits assessment tool user guide. Health Canada, Ottawa. [http://science.gc.ca/eic/site/063.nsf/eng/h\\_97170.html](http://science.gc.ca/eic/site/063.nsf/eng/h_97170.html).
36. AirQ+: software tool for health risk assessment of air pollution. [www.euro.who.int/en/health-topics/environment-and-health/air-quality/activities/airq-software-tool-for-health-risk-assessment-of-air-pollution](http://www.euro.who.int/en/health-topics/environment-and-health/air-quality/activities/airq-software-tool-for-health-risk-assessment-of-air-pollution).
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(5):381-9. doi: 10.1289/ehp.1307823.
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.* 11;49(1). pii: 1600419. doi: 10.1183/13993003.00419-2016.
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. doi: 10.1016/j.envint.2017.08.020.
40. Bob Carpenter et al., (2017). Stan: A probabilistic programming language. *Journal of Statistical Software* 76(1). DOI [10.18637/jss.v076.i01](https://doi.org/10.18637/jss.v076.i01).
41. R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
42. Samuelsen SO, Eide GE (2008). Attributable fractions with survival data. *Statistics in Medicine* 27:1447-1467.
43. Institute for Health Metrics and Evaluation (IHME). GBD Compare Data Visualization. Seattle, WA: IHME, University of Washington, 2016. Available from <http://vizhub.healthdata.org/gbd-compare>. (Accessed: March 13, 2017).



**Table S1.** Hazard ratios (10 µg/m<sup>3</sup>) with 95% confidence limits (in parentheses) and exposure ranges used in relative risk function estimation.

Cohort <sup>(ref #)</sup>	Region	Hazard Ratio <sup>+</sup>	Exposure Limits (µg/m <sup>3</sup> ) <sup>**</sup>	
			Lower	Upper
Breast <sup>(14)*</sup>	Canada	1.120 (1.047-1.198)	3.9	15.8
CanCHEC2001 <sup>(17)*</sup>	Canada	1.160 (1.130-1.190)	2.7	13.6
CanCHEC1991 <sup>(18)*</sup>	Canada	1.100 (1.080-1.120)	2.9	15.7
CCHS <sup>(19)*</sup>	Canada	1.260 (1.187-1.337)	2.4	12.9
ACS <sup>(11)*</sup>	US	1.070 (1.055-1.085)	6.7	20.9
AARP <sup>(12)*</sup>	US	1.030 (1.005-1.055)	6.9	22.7
NHS <sup>(15)*</sup>	US	1.130 (1.048-1.218)	5.8	20.5
CTS <sup>(16)*</sup>	US	1.010 (0.943-1.082)	3.4	29.4
SCS <sup>(26)</sup>	US	1.140 (1.068-1.217)	10.2	23.6
MHP <sup>(27)</sup>	US	0.860 (0.723-1.024)	12.3	22.4
AHS <sup>(28)</sup>	US	0.940 (0.781-1.131)	7.3	15.0
AHSMOG <sup>(29)</sup>	US	1.080 (0.967-1.206)	14.1	49.7
MEDICARE <sup>(32)</sup>	US	1.084 (1.081-1.086)	6.2	15.6
NHIS <sup>(22)*</sup>	US	1.060 (1.011-1.111)	8.1	17.3
China <sup>(10)*</sup>	Asia	1.064 (1.017-1.115)	15.4	83.7
Taipei <sup>(31)</sup>	Asia	0.920 (0.720-1.170)	24.9	32.3
Hong Kong <sup>(21)*</sup>	Asia	1.140 (1.068-1.217)	30.4	41.9
English <sup>(13)*</sup>	Europe	1.130 (1.003-1.273)	9.6	16.6
Dutch <sup>(29)</sup>	Europe	1.060 (0.969-1.159)	24.8	31.8
Rome <sup>(20)*</sup>	Europe	1.040 (1.030-1.050)	15.4	30.6
France <sup>(33)</sup>	Europe	1.150 (0.976-1.355)	9.9	24.1
DUELS <sup>(23)*</sup>	Europe	1.130 (1.110-1.150)	15.4	22.2

FINRISK <sup>(24)</sup>	Europe	0.771 (0.383-1.552)	5.9	9.4
HUBRO <sup>(24)</sup>	Europe	1.091 (0.699-1.703)	6.8	11.0
SNAC K <sup>(24)</sup>	Europe	1.086 (0.497-2.370)	5.8	10.2
SALT <sup>(24)</sup>	Europe	1.437 (0.706-2.927)	5.2	9.4
60 YR <sup>(24)</sup>	Europe	1.440 (0.579-3.580)	6.2	10.4
SDPP <sup>(24)</sup>	Europe	0.865 (0.262-2.861)	4.6	8.6
DCH <sup>(24)</sup>	Europe	1.106 (0.699-1.748)	9.8	12.8
EPIC MORGEN <sup>(24)</sup>	Europe	2.287 (0.745-7.022)	16.1	17.7
EPIC PROSPECT <sup>(24)</sup>	Europe	1.456 (0.526-4.030)	16.0	17.6
SALIA <sup>(24)</sup>	Europe	1.055 (0.564-4.030)	15.7	20.3
EPIC OXFORD <sup>(24)</sup>	Europe	1.020 (0.639-1.628)	8.0	11.6
KORA <sup>(24)</sup>	Europe	1.183 (0.510-2.740)	12.1	15.1
VHM &PP <sup>(24)*</sup>	Europe	1.151 (1.000-1.326)	10.4	16.1
SAPALDIA <sup>(24)</sup>	Europe	1.826 (0.365-9.134)	14.7	19.9
E3N <sup>(24)</sup>	Europe	1.238 (0.789-1.942)	11.9	18.1
EPIC TURIN <sup>(24)</sup>	Europe	1.777 (0.883-3.573)	26.8	33.4
SIDRIA TURIN <sup>(24)</sup>	Europe	1.757 (0.629-4.905)	28.2	33.8
SIDRIA ROME <sup>(24)</sup>	Europe	0.618 (0.292-1.310)	16.4	22.4
EPIC ATHENS <sup>(24)</sup>	Europe	1.044 (0.657-1.658)	16.0	24.8

\*: cohort with subject level PM<sub>2.5</sub>-mortality analysis

+: Hazard ratio and confidence limits as reported in literature.

\*\*.: Lower/Upper exposure limit for cohorts with subject-level analysis, indicated by \*, is their 1<sup>st</sup>/99<sup>th</sup> percentiles. Lower/Upper exposure limit for cohorts without subject-level analysis is their 5<sup>th</sup>/95<sup>th</sup> percentile.

**Table S2:** Global Exposure Mortality Model (GEMM) parameter estimates by cause of death, with and without inclusion of Chinese Male Cohort (10).

Cause of Death	Age Range (years)	With Chinese Male Cohort (10)					Without Chinese Male Cohort (10)				
		$\theta$	standard error $\theta$	$\alpha$	$\mu$	$\nu$	$\theta$	standard error $\theta$	$\alpha$	$\mu$	$\nu$
NCD+LRI	>25	0.1430	0.01807	1.6	15.5	36.8	0.1231	0.01797	1.5	10.4	25.9
	27.5	0.1585	0.01477	1.6	15.5	36.8	0.1358	0.01326	1.5	10.4	25.9
	32.5	0.1577	0.01470	1.6	15.5	36.8	0.1353	0.01321	1.5	10.4	25.9
	37.5	0.1570	0.01463	1.6	15.5	36.8	0.1348	0.01315	1.5	10.4	25.9
	42.5	0.1558	0.01450	1.6	15.5	36.8	0.1338	0.01304	1.5	10.4	25.9
	47.5	0.1532	0.01425	1.6	15.5	36.8	0.1317	0.01283	1.5	10.4	25.9
	52.5	0.1499	0.01394	1.6	15.5	36.8	0.1288	0.01255	1.5	10.4	25.9
	57.5	0.1462	0.01361	1.6	15.5	36.8	0.1256	0.01225	1.5	10.4	25.9
	62.5	0.1421	0.01325	1.6	15.5	36.8	0.1221	0.01194	1.5	10.4	25.9
	67.5	0.1374	0.01284	1.6	15.5	36.8	0.1181	0.01157	1.5	10.4	25.9
	72.5	0.1319	0.01234	1.6	15.5	36.8	0.1133	0.01112	1.5	10.4	25.9
	77.5	0.1253	0.01174	1.6	15.5	36.8	0.1077	0.01058	1.5	10.4	25.9
	85	0.1141	0.01071	1.6	15.5	36.8	0.0979	0.00964	1.5	10.4	25.9
IHD	>25	0.2969	0.01787	1.9	12	40.2	0.2543	0.04589	4.9	-21.1	17.7
	27.5	0.5070	0.02458	1.9	12	40.2	0.3996	0.03016	4.9	-21.1	17.7
	32.5	0.4762	0.02309	1.9	12	40.2	0.3796	0.02834	4.9	-21.1	17.7
	37.5	0.4455	0.02160	1.9	12	40.2	0.3512	0.02651	4.9	-21.1	17.7
	42.5	0.4148	0.02011	1.9	12	40.2	0.327	0.02468	4.9	-21.1	17.7
	47.5	0.3841	0.01862	1.9	12	40.2	0.3027	0.02285	4.9	-21.1	17.7
	52.5	0.3533	0.01713	1.9	12	40.2	0.2785	0.02103	4.9	-21.1	17.7
	57.5	0.3226	0.01564	1.9	12	40.2	0.2543	0.0192	4.9	-21.1	17.7
	62.5	0.2919	0.01415	1.9	12	40.2	0.2301	0.01737	4.9	-21.1	17.7
	67.5	0.2612	0.01266	1.9	12	40.2	0.2059	0.01554	4.9	-21.1	17.7
	72.5	0.2304	0.01117	1.9	12	40.2	0.1816	0.01371	4.9	-21.1	17.7
	77.5	0.1997	0.00968	1.9	12	40.2	0.1574	0.01188	4.9	-21.1	17.7
	85	0.1536	0.00745	1.9	12	40.2	0.1211	0.00914	4.9	-21.1	17.7
Stroke	>25	0.2720	0.07697	6.2	16.7	23.7	0.1873	0.08431	6.2	14.5	14.4
	27.5	0.4513	0.11919	6.2	16.7	23.7	0.3177	0.11625	6.2	14.5	14.4
	32.5	0.4240	0.11197	6.2	16.7	23.7	0.2985	0.1092	6.2	14.5	14.4
	37.5	0.3966	0.10475	6.2	16.7	23.7	0.2792	0.10216	6.2	14.5	14.4
	42.5	0.3693	0.09752	6.2	16.7	23.7	0.26	0.09511	6.2	14.5	14.4
	47.5	0.3419	0.09030	6.2	16.7	23.7	0.2407	0.08807	6.2	14.5	14.4
	52.5	0.3146	0.08307	6.2	16.7	23.7	0.2214	0.08102	6.2	14.5	14.4
	57.5	0.2872	0.07585	6.2	16.7	23.7	0.2011	0.07398	6.2	14.5	14.4
	62.5	0.2598	0.06863	6.2	16.7	23.7	0.1829	0.06693	6.2	14.5	14.4
	67.5	0.2325	0.06190	6.2	16.7	23.7	0.1637	0.05988	6.2	14.5	14.4
	72.5	0.2051	0.05418	6.2	16.7	23.7	0.1444	0.05284	6.2	14.5	14.4
	77.5	0.1778	0.04695	6.2	16.7	23.7	0.1252	0.0458	6.2	14.5	14.4
	85	0.1368	0.03611	6.2	16.7	23.7	0.0963	0.03523	6.2	14.5	14.4

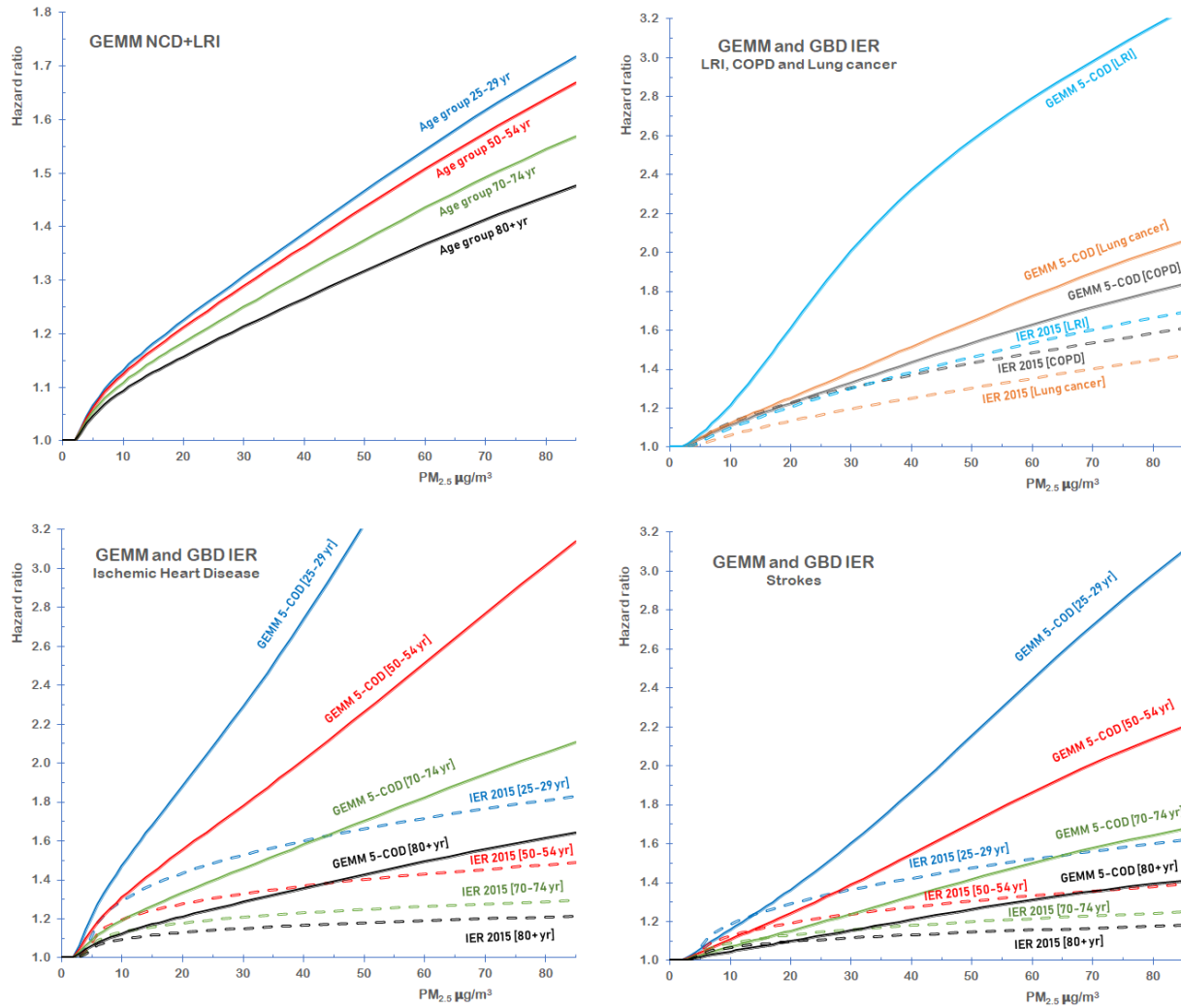
COPD	>25	0.2510	0.06762	6.5	2.5	32	0.2095	0.06725	7.2	2	14.7
Lung Cancer	>25	0.2942	0.06147	6.2	9.3	29.8	0.2626	0.07849	6.7	11	16.5
LRI	>25	0.4468	0.11735	6.4	5.7	8.4	NA	NA	NA	NA	NA

\*:  $GEMM(z) = \exp\{\theta \log(z/\alpha + 1) / (1 + \exp\{-(z-\mu)/\nu\})\}$ , where  $z = \max(0, PM_{2.5} - 2.4 \mu g/m^3)$

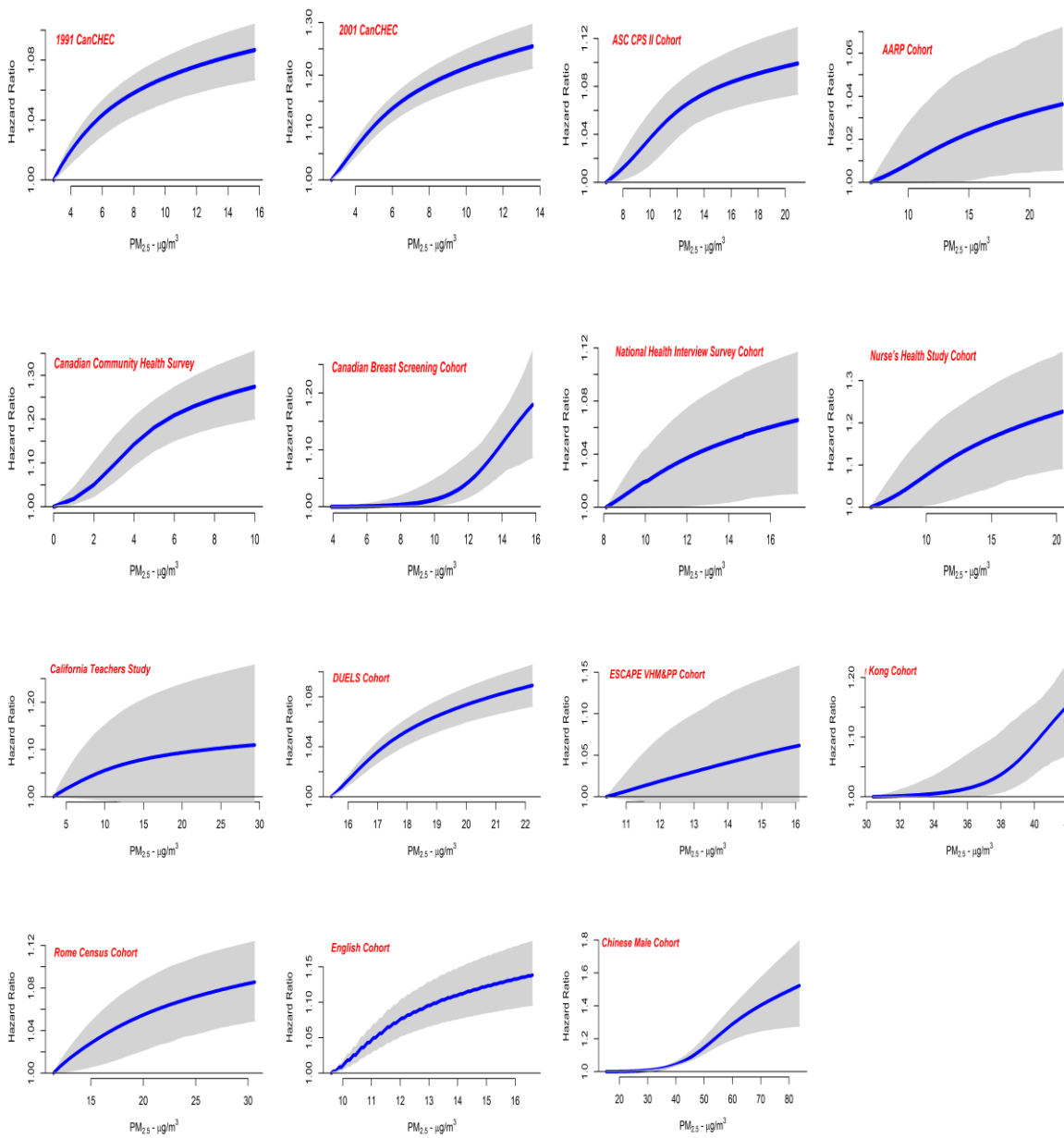
**Table S3:** Sensitivity of Global Exposure Mortality Model parameter estimates ( $\theta$ ,  $\alpha$ ,  $\mu$ ,  $\nu$ ) to variance and covariance specification of between cohort hazard ratio prediction error structures.

Error Structure	Variance Specification*	Covariance Specification*	$\theta$	Standard Error $\theta$	$\alpha$	$\mu$	$\nu$
Compound Symmetry	$\epsilon^2$	$\rho\epsilon^2$	0.1430	0.01807	1.6	15.5	36.8
Heteroscedastic Compound Symmetry	$\epsilon_s^2$	$\rho\epsilon_s^2\epsilon_l^2$	0.1432	0.01771	1.6	15.5	36.8
Unstructured	$\epsilon_s^2$	$\rho_{sl}\epsilon_s^2\epsilon_l^2$	0.1427	0.01781	1.6	15.5	36.7

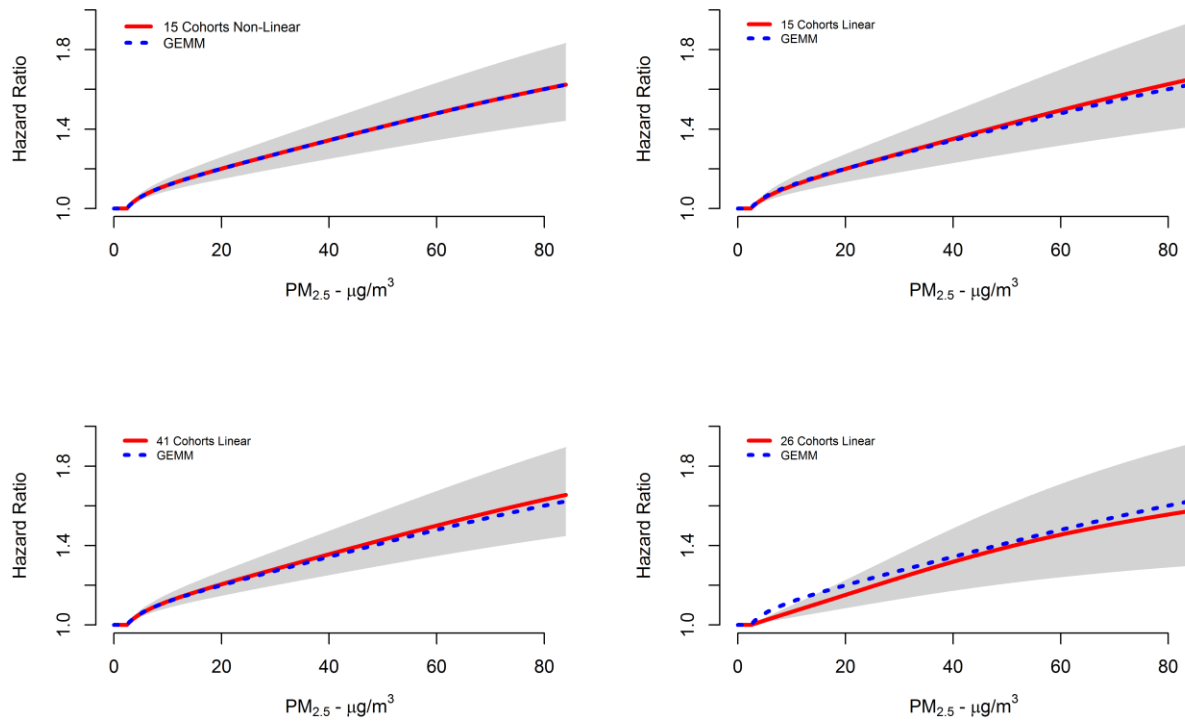
\*:  $s$  and  $l$  index study



**Fig S1** Hazard ratio predictions by  $PM_{2.5}$  concentration for the Global Exposure Mortality Model Non-Communicable Diseases Plus Lower Respiratory Infections (GEMM NCD+LRI) by selected age groups (upper left hand panel), GEMM and Integrated Exposure-Response (IER) for LRI, Chronic Obstructive Pulmonary Disease (COPD), and Lung Cancer (upper right hand panel), GEMM and IER for Ischemic Heart Disease (IHD) by selected age groups (lower left hand panel), and for Stroke (lower right hand panel).

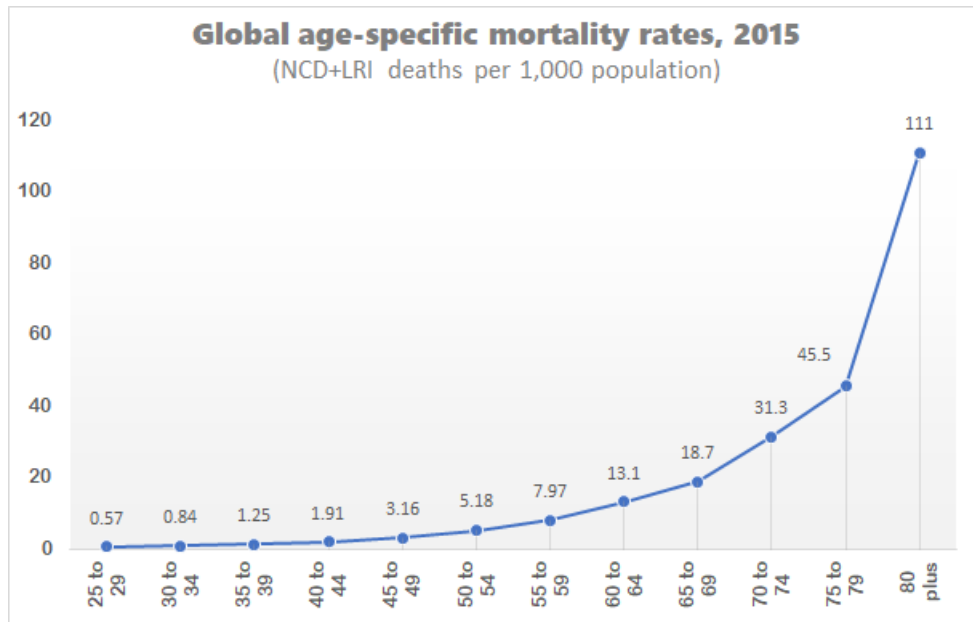


**Fig S2** Fine Particulate Matter-Non-Accidental Mortality association (blue line) and 95% confidence intervals (grey shaded area) by cohort. Cohort-specific relative risks are calculated as a contrast from concentrations within the cohort-specific range to the lowest concentration of that cohort. Ensemble estimate and confidence intervals presented for the Chinese Male Cohort (10) for two models involving a model that included individual level risk factors such as smoking, obesity, and diet and another model that included individual plus contextual risk factors, such as urban/rural, region, and percent low education. Ensemble model assumed equal weights among these two models.

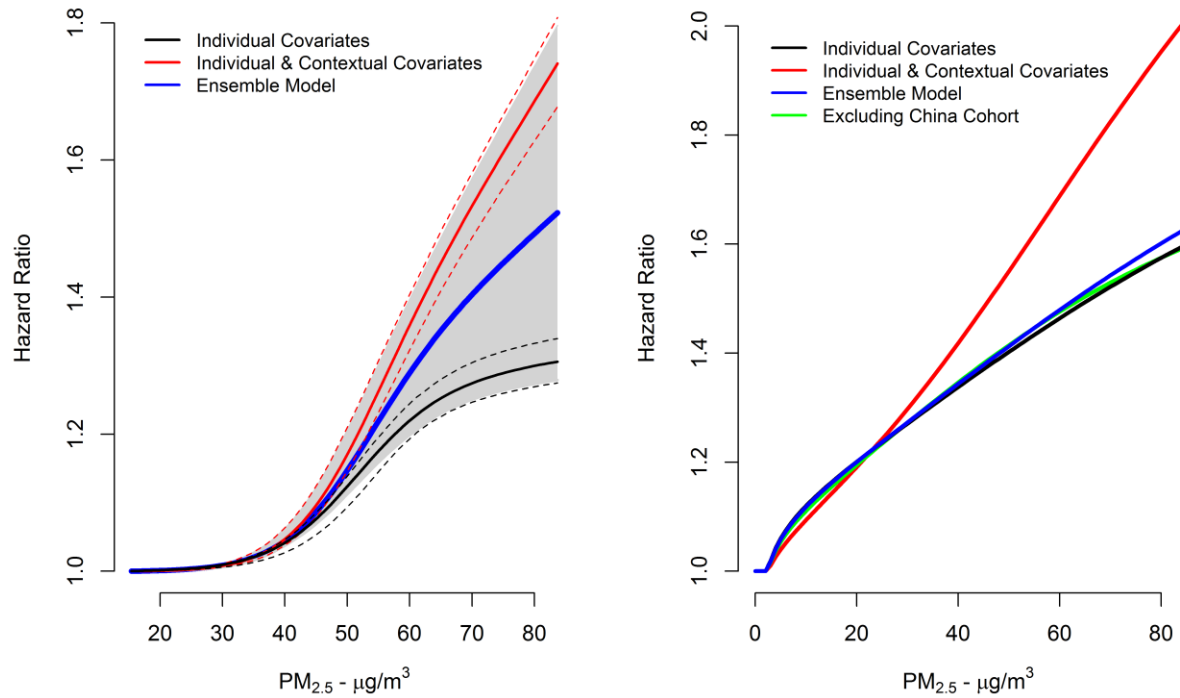


**Fig S3** Comparison of the Global Exposure Mortality Model (GEMM) based on 15 cohorts with non-linear within cohort PM<sub>2.5</sub>-mortality associations and 26 cohorts assuming a linear association (blue dashed line) to alternative specifications (mean prediction - red solid line; 95% uncertainty interval – grey shaded area): upper left panel – using only 15 cohorts with non-linear within cohort association; upper right hand panel – using only 15 cohorts assuming a linear within cohort association; lower left hand panel – using all 41 cohorts assuming a within cohort linear association; and lower right hand panel – using 26 cohorts assuming a linear within cohort association.

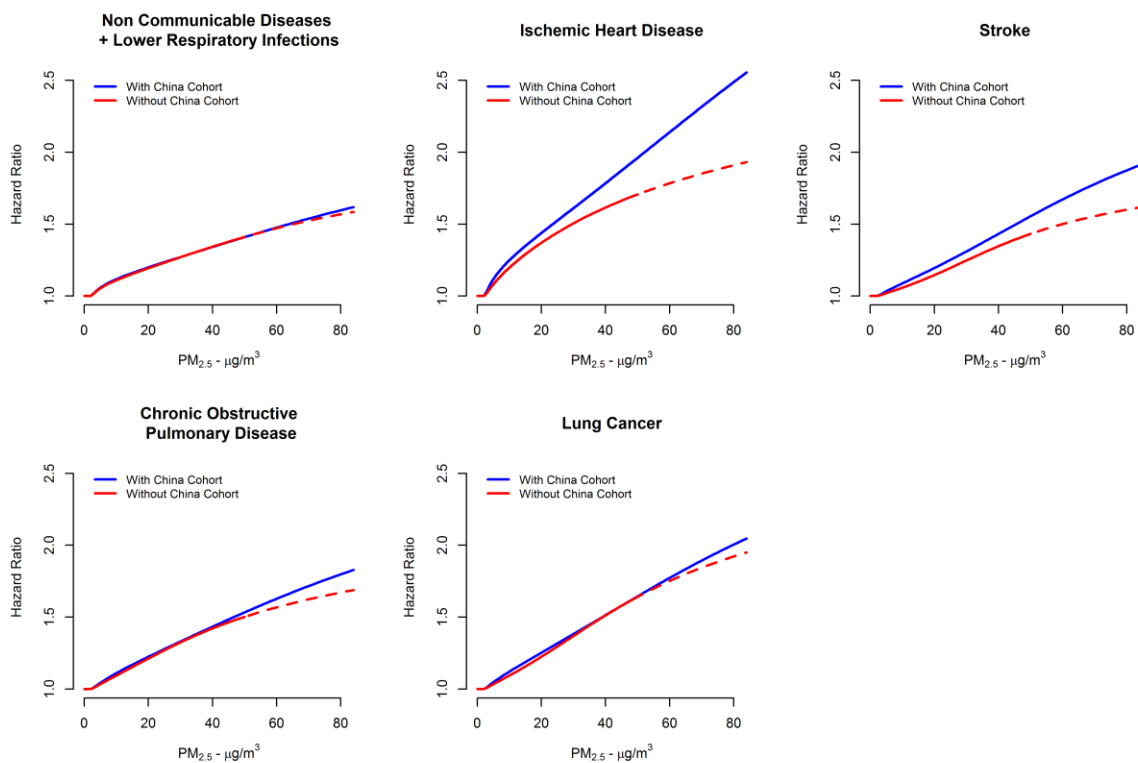




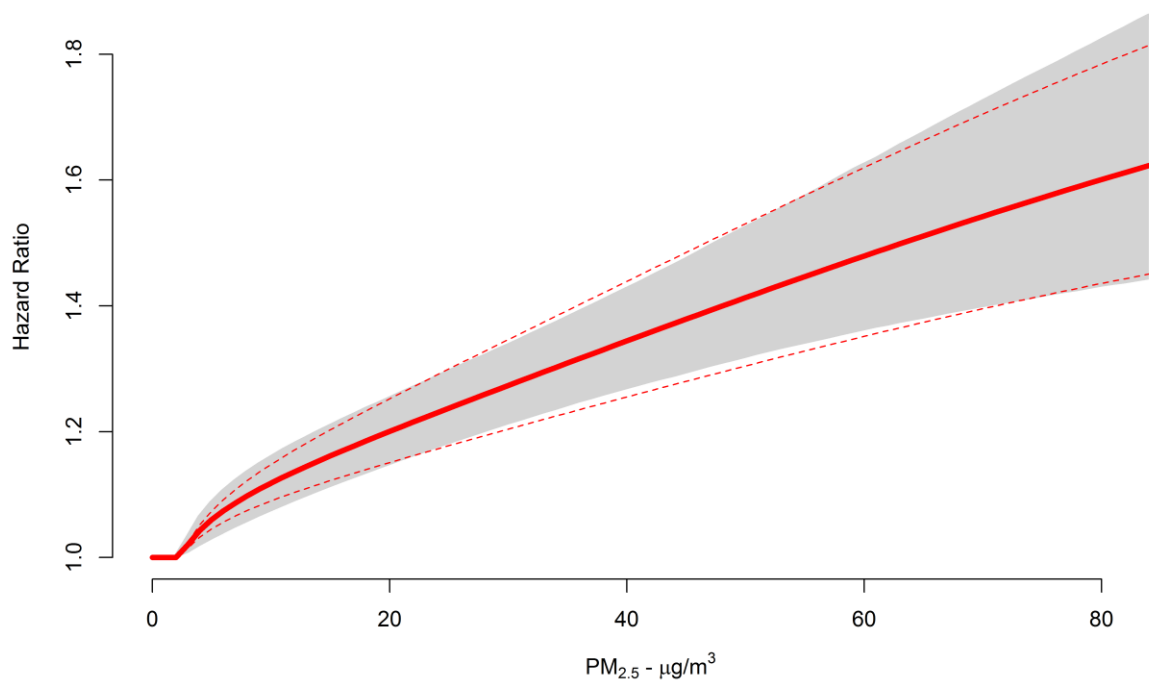
**Fig S4** Global Non-Communicable Diseases plus Lower Respiratory Infection (NCD+LRI) mortality rates per 1,000 population by age group (43).



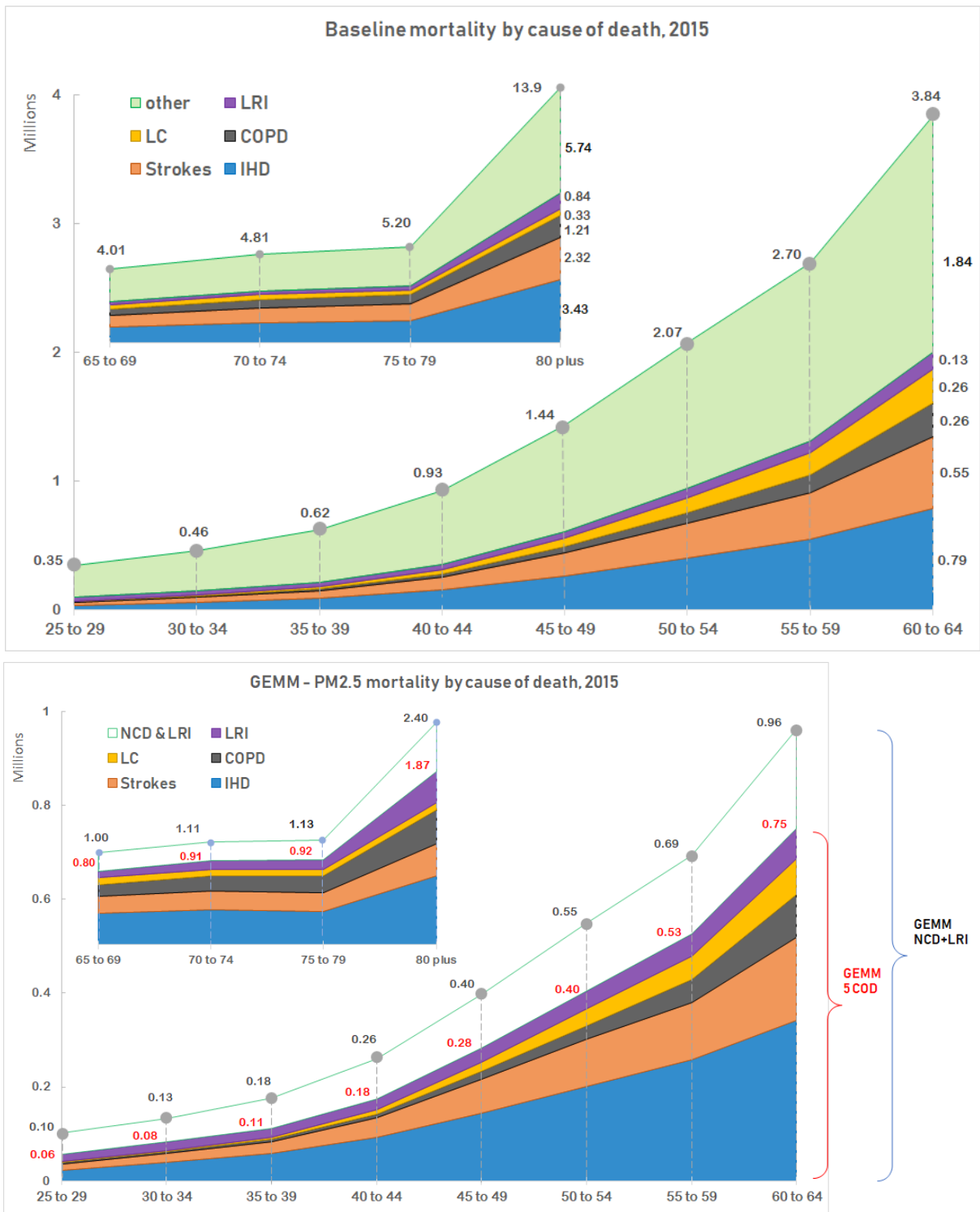
**Fig S5:** Influence of the hazard ratio model selected for the Chinese Male Cohort (10) on the predicted values of the resulting Global Exposure Mortality Models (GEMM). Models including only individual covariates, such as smoking and diet (black line of left hand panel), models including both individual and contextual covariates, such as urban/rural indicator, region of China, and percentage low education (red line of left hand panel), and ensemble model assuming equal weights among these two models (blue line left hand panel). Resulting *GEMM* predictions presented in right hand panel. Additional *GEMM* excluding China cohort also presented (green line of right hand panel).



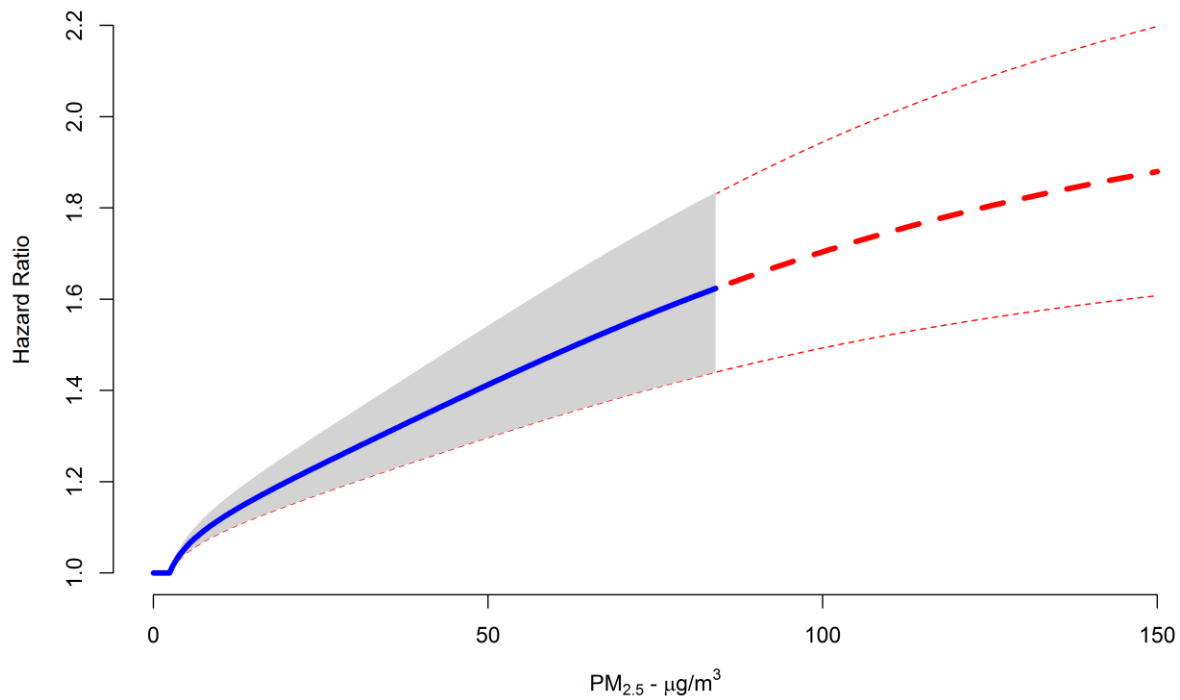
**Fig S6** Global Exposure Mortality Model (GEMM) hazard ratio predictions over the PM<sub>2.5</sub> exposure range for Non-Communicable Diseases Plus Lower Respiratory Infections (NCD+LRI), Ischemic Heart Disease (IHD), Stroke, Chronic Obstructive Pulmonary Disease (COPD), and Lung Cancer with (blue line) and without Chinese Male Cohort (red line). Extrapolated predictions beyond the highest concentration without the Chinese cohort (50µg/m<sup>3</sup>) are also presented (dashed red line). NCD+LRI, IHD, and Stroke GEMMs for the 60-64 year old age group.



**Fig S7** Global Exposure Mortality Model hazard ratio predictions over PM<sub>2.5</sub> exposure range for ensemble model (blue solid line) with 95% uncertainty interval (grey shaded area) and approximation to ensemble (red solid line) and 95% uncertainty interval (dashed red lines). Note ensemble model and its approximation are almost identical and thus blue line does not show.



**Fig S8** Global numbers of deaths (millions) by age group and cause (upper panel). Global numbers of deaths attributed to PM<sub>2.5</sub> exposure by age group and cause based on NCD+LRI GEMM and 5 COD GEMM.



**Fig S9** Global Exposure Mortality Model (GEMM) predictions over observed concentration range (2.4 µg/m<sup>3</sup> to 84 µg/m<sup>3</sup>) (solid blue line) and 95% uncertainty interval (grey shaded area). Extrapolation beyond range of exposure (84 µg/m<sup>3</sup> to 150 µg/m<sup>3</sup>) (dashed red line) and 95% uncertainty interval (dotted red line).