Supplementary Appendix for

National and county life expectancy loss due to particulate matter pollution in the USA

James E. Bennett[†], Helen Tamura-Wicks[†], Robbie M. Parks, Richard T. Burnett, C. Arden Pope III, Matthew J. Bechle, Julian D. Marshall, Goodarz Danaei, Majid Ezzati

† Equal contribution, listed alphabetically

Table of Contents

Study design

Our aim was to quantify, nationally as well as at the small-area (county) level, the current mortality burden of fine particulate matter ($PM_{2.5}$) pollution in the contiguous USA (i.e., excluding Alaska and Hawaii) and the averted mortality as a result of recent $PM_{2.5}$ reductions.

The most common source of information on the health effects of air pollution exposure is from prospective cohort studies [1-5]. By design, prospective cohort studies provide information on the magnitude of the risk associated with $PM_{2.5}$ in a specific group of participants after a follow-up period, during which other determinants of population health also change. Therefore, estimates of excess mortality and loss of life expectancy due to air pollution in the contemporary population [5,6], and the benefits of control, typically rely on applying risk estimates from prospective cohorts to current $PM_{2.5}$ concentrations and mortality statistics, with an implicit assumption that the magnitude of risk associated with $PM_{2.5}$ in the prospective cohort participants is the same as the contemporary general population.

A direct approach for measuring the mortality burden of air pollution is to directly analyze the association of death rates (or life expectancy as a summary measure) with $PM_{2.5}$ concentration across a large number of "small-area" units, such as counties, appropriately adjusting for other determinants of mortality [7-11]. Here, we used four Bayesian spatiotemporal models, with different degrees of adjustment for county characteristics that may also affect mortality, to directly estimate excess deaths and loss of life expectancy from $PM_{2.5}$ in US counties from 1999 to 2015. We began the analysis in 1999 for two reasons: First, reliable data on $PM_{2.5}$ throughout the USA became available in 1999, when the US Environmental Protection Agency set up the Federal Reference Method network of $PM_{2.5}$ monitors [12]. Second, the USA switched from the 9th revision to the 10th

revision of the International Classification of Disease (ICD) system in 1999 [13]. Therefore, the post-1999 data are based on a consistent system of the assignment of the medical cause of death.

Statistical methods

We formulated and implemented four Bayesian spatio-temporal models to directly estimate the effect of $PM_{2.5}$ concentration on cardiorespiratory diseases death rates at the county level. All analyses were done separately by sex and age group (five year age groups from birth to 85 years, and 85 years and older) because death rates vary by age group and sex, as might their associations with air pollution [14].

The four models had similar structures, but differed in the extent of adjustment for county characteristics that may also influence mortality, ranging from unadjusted to restrictive as described below and shown in Table S1. Specifically, for a given county, sex, and age group, the number of deaths in each year follows a Poisson distribution:

$$deaths_{county-year} \sim Poisson(death\ rate_{county-year} \cdot population_{county-year})$$

In all four models, log-transformed death rates were modelled as a sum of components that depend on time and air pollution, with some models also including terms for other time-varying county characteristics.

All models contained terms to capture the overall level and rate of change of mortality, with α_0 as the common intercept for log-transformed death rates, and β_0 the common time slope. Non-linear trends were captured by a first-order random walk ν_{year} [15]. All models also included a term that relates log-transformed death rate to county $PM_{2.5}$ concentration, with the coefficient γ representing the logarithm of the rate ratio per 1 μg/m^3 higher concentration. $\varepsilon_{county-year}$ is an over-dispersion term that captures the variation unaccounted for by other terms in the model, and is modelled as $N(0, \sigma_{\varepsilon}^{2})$.

The *unadjusted* model did not include any terms related to county characteristics. The *covariate* model included terms for seven time-varying county level covariates $X_{i-county-year}$ (i=1…7) (per capita income; percentage of population whose family income is below the poverty threshold, who are of Black race, who have graduated from high-school, who live in urban areas, and who are unemployed; and a proxy for cumulative smoking) to account for other potential determinants of mortality with associated covariate-specific slopes, θ_i. This model is equivalent to the approach used by cohort studies, which typically adjust for baseline characteristics of subjects.

The *covariate-and-county model* also included county-specific random intercepts (α_{county}), which largely removes systematic variations in death rates and $PM_{2.5}$ concentration across counties, hence implicitly adjusting for unobserved characteristics that systematically lead to higher or lower pollution and mortality in each county. In this model, the mortality effects of air pollution are largely inferred through cross-county differentials in changes in death rates and $PM_{2.5}$ concentrations over time. The *restrictive* model included both county-specific random intercepts and random slopes (β_{county}), hence also implicitly adjusting for linear differential change in death rates and $PM_{2.5}$ concentrations across countries and relates changes in $PM_{2.5}$ concentration to changes in death rates, after removing linear trends in mortality and pollution. By design, this model leaves little spatial or temporal variation in death rates to be associated with $PM_{2.5}$.

The county random intercepts and slopes were modelled using a conditional autoregressive (CAR) structure that (empirically) allows for death rates to be more similar across neighboring counties than those that are further away.

We fitted the models using integrated nested Laplace approximation (INLA), using the R-INLA software, which offers orders of magnitude of computational efficiency improvement in Bayesian inference compared to traditional MCMC for latent Gaussian models. As in previous work [16], weakly informative hyper-priors of logGamma(1, 0.001) were specified on the logarithm of the precisions of the random effects so that the parameter estimates were data driven.

We used the observed age-specific death rates to calculate life expectancy for each county or merged county unit, using standard life table techniques [17]. We used the Kannisto-Thatcher method to expand the terminal (85 years and older) age group of the life table [18]. This method is designed for use in low-mortality populations and is used by the UN Population Division and the World Health Organization [19].

We calculated life expectancy under the counterfactual scenario of alternative $PM_{2.5}$ concentration (e.g., the lowest observed concentration) by calculating the number of deaths from cardiorespiratory diseases if $PM_{2.5}$ concentrations were at their alternative level [20]. The difference between the observed and counterfactual life expectancy measures the contribution of $PM_{2.5}$ pollution to life expectancy loss, see in Fig. 2 and 3 in the main paper. For the contribution of $PM_{2.5}$ reduction to life expectancy gain from 1999 to 2015, we calculated 1999 life expectancies using observed death rates as well as those expected if $PM_{2.5}$ concentration in each county had been at its 2015 level.

Figure S1. Sex and age group specific rate ratios per 10 $\mu g/m^3$ of $PM_{2.5}$ for cardiorespiratory deaths. The rate rations were estimated from four different Bayesian spatio-temporal models described in Statistical Methods.

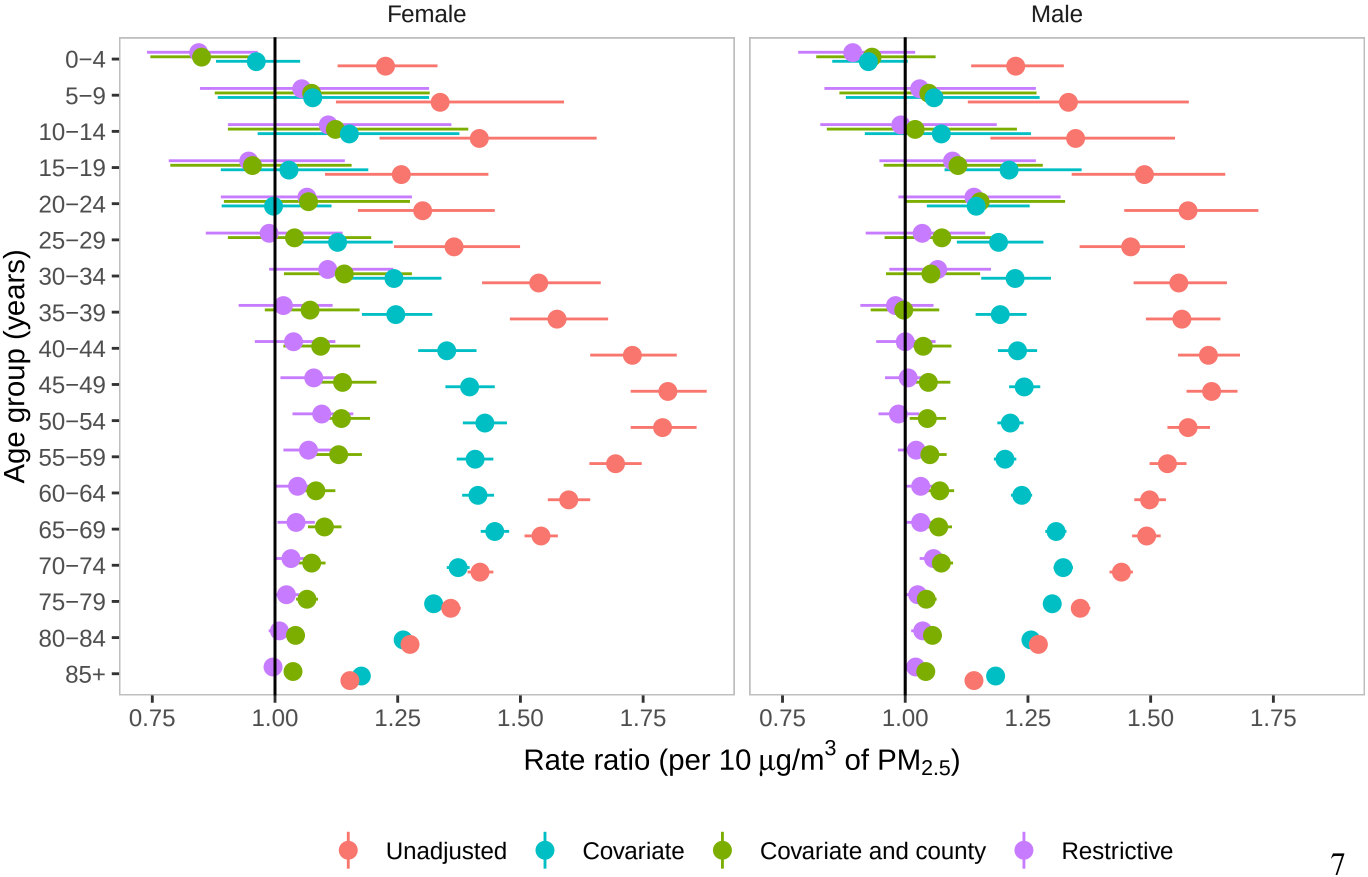
Female
Male
Age group (years)
0−4
5−9
10−14
15−19
20−24
25−29
30−34
35−39
40−44
45−49
50−54
55−59
60−64
65−69
70−74
75−79
80−84
85+
0.75
1.00
1.25
1.50
1.75
Rate ratio (per 10 μg/m3 of PM2.5)
Unadjusted
Covariate
Covariate and county
Restrictive

Table S1. Models used for relating county death rates to $PM_{2.5}$ concentrations.

Model		Common terms	County terms	Pollution terms	Covariates	Over-dispersion
Unadjusted	$\log(death\ rates_{county-year}) =$	$\alpha_0 + \beta_0 \cdot year + \nu_{year}$		$+ \ \gamma \cdot PM_{county-year}$		$+ \ \varepsilon_{county-year}$
Covariate	$\log(death\ rates_{county-year}) =$	$\alpha_0 + \beta_0 \cdot year + \nu_{year}$		$+ \ \gamma \cdot PM_{county-year}$	$+ \ \sum_{i=1}^{i=7} \theta_i \cdot X_{i-county-year}$	$+ \ \varepsilon_{county-year}$
Covariate-and-county	$\log(death\ rates_{county-year}) =$	$\alpha_0 + \beta_0 \cdot year + \nu_{year}$	$+ \ \alpha_{ccounty}$	$+ \ \gamma \cdot PM_{county-year}$	$+ \ \sum_{i=1}^{i=7} \theta_i \cdot X_{i-county-year}$	$+ \ \varepsilon_{county-year}$
Restrictive	$\log(death\ rates_{county-year}) =$	$\alpha_0 + \beta_0 \cdot year + \nu_{year}$	$+ \ \alpha_{ccounty} + \beta_{county} \cdot year$	$+ \ \gamma \cdot PM_{county-year}$	$+ \ \sum_{i=1}^{i=7} \theta_i \cdot X_{i-county-year}$	$+ \ \varepsilon_{county-year}$

References

1. Dockery DW, Pope CA, Xu XP, et al. An association between air-pollution and mortality in 6 United-States cities. New Engl J Med 1993;329:1753-9.

2. Pope CA, 3rd, Burnett RT, Thun MJ, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 2002;287:1132-41.

3. Pope CA, Dockery DW. Health effects of fine particulate air pollution: lines that connect. J Air Waste Manage 2006;56:709-42.

4. Brook RD, Rajagopalan S, Pope CA, et al. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. Circulation 2010;121:2331-78.

5. Burnett R, Chen H, Szyszkowicz M, et al. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. P Natl Acad Sci USA 2018;115:9592-7.

6. Fann N, Kim SY, Olives C, Sheppard L. Estimated changes in life expectancy and adult mortality resulting from declining PM2.5 exposures in the contiguous United States: 1980-2010. Environ Health Perspect 2017;125:097003.

7. Pope CA, Ezzati M, Dockery DW. Fine-particulate air pollution and life expectancy in the United States. New Engl J Med 2009;360:376-86.

8. Correia AW, Pope CA, 3rd, Dockery DW, Wang Y, Ezzati M, Dominici F. Effect of air pollution control on life expectancy in the United States: an analysis of 545 U.S. counties for the period from 2000 to 2007. Epidemiology 2013;24:23-31.

9. Elliott P, Wakefield JC, Best NG. Spatial Epidemiology: methods and applications: Oxford University Press; 2001.

10. Lave LB, Seskin EP. Air pollution and human health. Science 1970;169:723.

11. Özkaynak H, Thurston GD. Associations between 1980 US mortality rates and alternative measures of airborne particle concentration. 1987;7:449-61.

12. Kim S-Y, Olives C, Sheppard L, et al. Historical prediction modeling approach for estimating long-term concentrations of PM2.5 in cohort studies before the 1999 implementation of widespread monitoring. Environ Health Persp 2017;125:38-46.

13. International Classification of Diseases 10th Revision (ICD-10). at https://www.cdc.gov/nchs/data/dvs/icd10fct.pdf.)

14. Crouse DL, Peters PA, Hystad P, et al. Ambient PM2. 5, O3, and NO2 exposures and associations with mortality over 16 years of follow-up in the Canadian Census Health and Environment Cohort (CanCHEC). Environ Health Persp 2015;123:1180.

15. Rue H, Held L. Gaussian Markov random fields: theory and applications: CRC press; 2005.

16. Kontis V, Bennett JE, Mathers CD, Li GQ, Foreman K, Ezzati M. Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble. Lancet 2017;389:1323-35.

17. Preston S, Heuveline P, Guillot M. Demography: measuring and modeling population processes2000.

18. Thatcher AR, Kannisto V, Vaupel JW. The force of mortality at ages 80 to 120. 1998.

19. WHO methods and data sources for life tables 1990-2016. 2018. at http://www.who.int/healthinfo/statistics/LT_method.pdf.)

20. Danaei G, Rimm EB, Oza S, Kulkarni SC, Murray CJL, Ezzati M. The promise of prevention: the effects of four preventable risk factors on national life expectancy and life expectancy disparities by race and county in the United States. Plos Med 2010;7.