Dependence of Short-Term Mortality on Fine Particulate Matter in the Population of Elderly Medicare Beneficiaries

Richard L. Smith
Department of Statistics and Operations Research,
University of North Carolina, Chapel Hill

November 17, 2021

Abstract

Purpose: To investigate the sensitivity of short-term associations between mortality in the Medicare population and fine particulate matter (PM) to various statistical modeling assumptions. Methods: Mortality data were downloaded from Medicare, particulate matter data from EPA, temperature and dewpoint from NOAA. The case-crossover method was used to evaluate the association between mortality and PM (average of lags 0 and 1 day) with comparison days on the same day of week in fixed 28-day windows. Three concentration-response functions were considered: linear, nonlinear, and a "broken stick" model hinged at 12 μg/m^3. Nonlinear functions of temperature and dewpoint, both on day of death and average of lags 1–3 days, were also included. Sensitivity analyses included age, sex and region. Results: Significant associations were found when a linear concentration response function was fitted to the full range of PM, or in a broken-stick model above 12 μg/m^3. No significant association was found below 12 μg/m^3. However when lagged meteorology was omitted from the model, the estimated coefficients greatly increased and were significant at all levels of PM. Conclusions: It is important to take lagged meteorology into account in investigating short-term associations between PM and mortality.

1 Introduction

Over the past twenty years, there have been many papers summarizing the association between airborne concentrations of particulate matter and mortality or other adverse health effects in the

U.S. or worldwide population. Studies may be broadly divided into two categories: those examining short-term or acute effects, typically of duration under a week [26, 9, 23, 24, 27, 13, 8, 25, 20, 11, 33, 5, 32, 6], and those concerned with long-term effects [16, 34, 17, 2, 7, 1, 4]. Although both short-term and long-term effects are important in a regulatory context, the present paper is solely about short-term effects.

The United States Environmental Protection Agency (EPA) is mandated by the Clean Air Act to promulgate air pollution standards that are "requisite to protect the public health ... allowing an adequate margin of safety." These standards are reviewed every five years to determine whether they are sufficient to meet that requirement. The current EPA standard for particulate matter of aerodynamic diameter of 2.5 μm or less ($PM_{2.5}$) requires an annual average of 12 $\mu g/m^3$ or less and a daily maximum of 35 $\mu g/m^3$ or less in each community. In practice, the daily standard is only exceeded by a tiny minority of monitored $PM_{2.5}$ readings, whereas 12 $\mu g/m^3$ is near the median. Therefore, there is particular interest in understanding health effects at levels that are at or below the level of 12 $\mu g/m^3$. Evidence of a health effect in this range would strengthen the case for a tightening of the standard.

In this paper, we focus specifically on the short-term mortality effect. Recent studies [6, 10] have used the "case-crossover" method of statistical analysis and have suggested strong associations between $PM_{2.5}$ and short-term mortality even below 12 $\mu g/m^3$. The objectives of the present paper are to investigate:

1. Robustness of the results against alternative constructions of the meteorological and air pollution databases;

2. Robustness of the results against the inclusion of alternative lags of meteorology;

3. Alternative concentration-response functions for examining the $PM_{2.5}$–mortality association both above and below 12 $\mu g/m^3$;

4. Regional variations in the associations.

The last of these is motivated by previous research showing strong regional variations in the ozone–mortality association [3, 30], with effects in the north-east USA being generally stronger than elsewhere. There has been limited research whether similar regional effects also hold for $PM_{2.5}$

(e.g. [34]), and an earlier paper on short-term mortality using data for California [32] claimed effectively zero association between $PM_{2.5}$ and mortality over 2000-2012 in that state. Among other conclusions, the present paper confirms the result of [32], though all three major sources of data (mortality, meteorology and $PM_{2.5}$) have been constructed from entirely independent sources and the method of statistical analysis is also different.

2 Data and Statistical Methods

Medicare data were obtained from the Center for Medicare and Medicaid Services for the years 1999–2013. For each deceased individual, the date of death, zipcode of residence at time of death, age, sex and race were recorded. Meteorology data was obtained from the National Centers for Environmental Information. For each zipcode, daily temperature and dewpoint were recorded from the nearest weather station within 100 km. $PM_{2.5}$ data were derived from both EPA monitors and the EPA's Remote Sensing Information Gateway, which combines information from air quality models and monitors. Further details of both the data processing are given in Supplementary Materials, Section 1. The exposure variable in this study is defined as the average of $PM_{2.5}$ on day of death and the one day preceding. Temperature and dewpoint data are calculated on day of death and the mean of three days prior to death (lagged meteorology).

The statistical analysis in this paper is based on the case-crossover method [19, 18, 14, 15, 28]. Each date of death is matched with nearby "referent" dates whose air pollution and meteorology are compared with those of the day of death. Theoretical analyses have shown the importance of using predetermined "referent windows" and using all of the dates within a referent window to avoid biases created by missing data. In this analysis, each date of death is matched with three other days at seven-day intervals within a fixed window of length 28 days.

Subgroup analyses may be based on sex (separate analyses for male and female beneficiaries); by race; by age group; and by region of country: whole US, North-East, South-East, North-West, South-West and a separate analysis for California.

All analyses assume that the logged mortality rate is a function of temperature, dewpoint and $PM_{2.5}$. Temperature and dewpoint are both modeled nonlinearly through B-splines with varying degrees of freedom (DF) with a default DF=6. The models for $PM_{2.5}$ were (a) linear; (b) nonlinear modeled by B-splines; or (c) the "broken stick" formula

$$f(x) \quad = \quad \begin{cases} \beta_1(x-12) & \text{if } 12 < x <= 35, \\ \beta_2(x-12) & \text{if } x <= 12, \end{cases} \tag{1}$$

which represents the effect as two straight lines joined at $x = 12$. The rationale behind (1) is that the coefficients β_1 and β_2 represent the $PM_{2.5}$ effect over the two ranges that are of greatest regulatory interest: β_2 below the long-term standard, and β_1 between the long-term and daily standards. In this way, we hope to get a clear-cut numerical determination (with confidence limits) of the $PM_{2.5}$ effect over both of those ranges. All $PM_{2.5}$ coefficients are expressed in units of percent rise in mortality per 10 μg/m^3 rise in $PM_{2.5}$.

With monitored $PM_{2.5}$, we have 1,769,871 complete rows of data (no missing meteorology or $PM_{2.5}$ values on either the date of death or the three comparison days). This number rises to 16,196,012 using the downscaled $PM_{2.5}$ data over 2002–2013. Many $PM_{2.5}$ monitors only take readings every third day which limits their use for this kind of analysis.

Summary statistics for meteorology and $PM_{2.5}$ are contained in Table 1. This table shows the minimum and maximum values along with their quantiles at the 2.5%, 25%, 50%, 75% and 97.5% points of the distribution. These refer to day of death only and (except for the last line) the larger dataset based on downscaled $PM_{2.5}$. The rows for lagged temperature and lagged dewpoint refer to three-day averages of daily means prior to the day of death, and are computed separately as it can be expected that three-day averages would have a less dispersed distribution than single-day means. For both the monitored and downscaled $PM_{2.5}$ data, the calculations are for two-day averages, as these are the values used in our epidemiological models.

Summaries of the deaths are given in Table 2, classified by sex, race, age group and region. Overall, we have 45.9% male; 84.3% white; division among age groups is 8.4% (under 65); 20.1% (65–74); 33.5% (75–84); 38.0% (85+); division among regions is 41.3% (North-East); 23.7% (South-East); 9.7% (North-West); 25.4% (South-West). The total of all individuals in Table 2 is reduced slightly by the fact that not all individuals are classified by sex and race, and the classification into regions is restricted to those who live in the continental United States. Some additional summary tables are given in Supplementary Materials, Section 2.

One peculiarity of the data was identified in the initial data processing: it appears that deaths

Variable	Minimum	Q2.5	Q25	Q50	Q75	Q97.5	Maximum
Daily Mean Temperature	–37.3	20.3	44.5	59.4	71.9	85.1	109.9
Lagged Temperature	–23.8	21.5	44.7	59.2	71.6	84.6	109.2
Daily Mean Dewpoint	–45.9	7.8	31.0	46.1	59.2	73.3	86.7
Lagged Dewpoint	–31.3	10.2	31.4	45.6	58.7	73.0	83.7
$PM_{2.5}$ (downscaled)	0.4	3.9	7.1	9.8	13.7	26.0	125.9
$PM_{2.5}$ (from monitor)	0.0	3.8	7.8	11.2	16.2	33.1	170.1

Table 1: Summary statistics for meteorology and $PM_{2.5}$ data

Region	North-East		South-East		North-West		South-West		
Race	White	Other	White	Other	White	Other	White	Other	Totals
M, 0–64	220,942	85,171	147,791	76,134	61,631	9,569	144,607	61,747	807,592
F, 0–64	147,551	59,175	97,485	55,762	43,582	6,151	97,682	42,118	549,506
M, 65–74	568,187	132,750	362,730	103,372	153,027	12,388	381,291	98,365	1,812,110
F, 65–74	460,820	111,065	276,296	85,298	121,991	9,239	297,034	77,867	1,439,610
M, 75–84	942,349	140,404	524,946	102,349	238,628	13,379	554,919	126,568	2,643,542
F, 75–84	996,089	161,201	533,136	120,744	243,249	13,750	569,495	130,025	2,767,689
M, 85+	817,465	79,903	403,062	62,469	221,147	8,667	465,838	87,069	2,145,620
F, 85+	1,573,412	169,674	734,133	137,309	394,689	13,739	822,217	141,990	3,987,163
Totals	5,726,815	939,343	3,079,579	743,437	1,477,944	86,882	3,333,083	765,749	16,152,832

Table 2: Classification of Medicare deaths by sex, age group, region and race

on the last day of each month are substantially higher than on the remaining days of the month (see Supplementary Materials, Section 3). A reviewer has pointed out that this is a known feature of Medicare data and can be avoided by using a flag that identifies which dates of death are validated: however, this was only pointed out after the main analyses of the paper had been completed. Instead, the analyses used three strategies for correcting for this anomaly: (a) ignore it; (b) omit the last day of each month; (c) omit the last day of each month and any days matched with the last day of a month in the case-crossover analysis. Initial results and simulations showed that method (b) raises significant bias issues that are similar to the phenomenon of "overlap bias" [15], so the bulk of the following reported results used method (c).

The analysis used a Fortran program to input the data and compute the likelihood function; maximum likelihood estimation then proceeded using a variable metric algorithm[22]. The main limitation on this style of analysis was the number of observations that could be processed in memory; the largest number processed in any single analysis was about 11.3 million. The dataset consisting of all individuals aged 65 and over contains about 12.4 million individuals; this dataset was split into male and female and the results combined using a meta-analysis approach.

The bulk of the analyses used daily deaths from 2002-2013 with $PM_{2.5}$ data from RSIG; temperature and dewpoint were used either with day of death only (the No Lags model) or with day of death and the mean of the three previous days (the With Lags model). For analyses with a log-linear relationship between $PM_{2.5}$ and mortality, the range of $PM_{2.5}$ was unrestricted. For the broken stick analyses, the range was confined to 0–35 μg/m^3. Because of the extra computational cost of the broken stick analysis, not all analyses were conducted under both the log-linear and broken stick models. Unless reported otherwise, all analyses omit the last day of each month and any other days matched with the last day of a month under the case-crossover analysis. In addition to analyses for the whole USA, separate analyses were performed for the North-East, South-East, North-West, South-West and California. Some analyses divided by participants by race, classified here as either white or non-white (other races were merged into non-white).

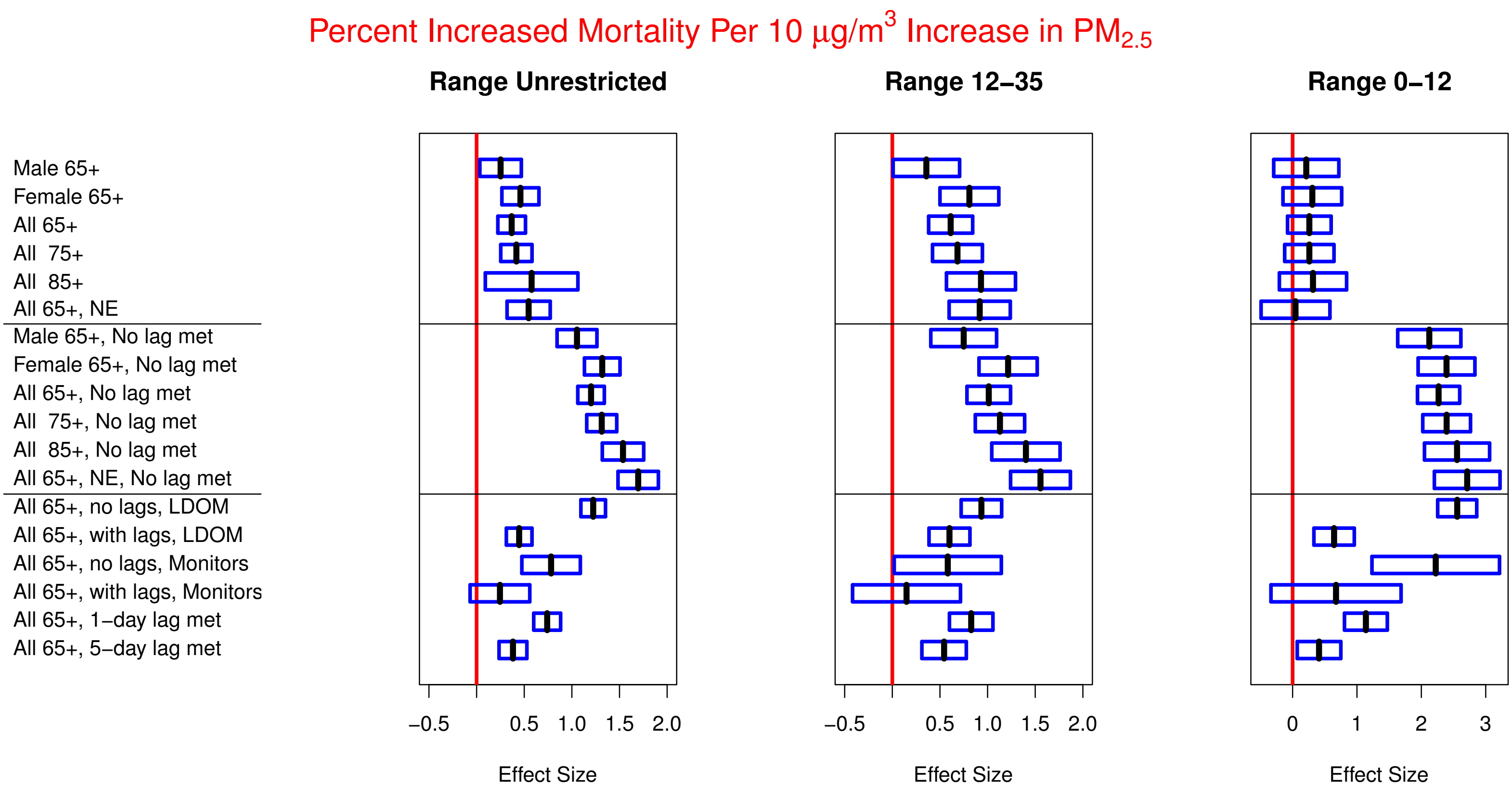

Figure 1: Estimated percent change in mortality and 95% confidence intervals associated with 10 μg/m^3 rise in $PM_{2.5}$ for various subpopulations and statistical models. Left group of plots: linear concentration-response function fitted to full range of $PM_{2.5}$. Middle and right groups: "broken stick" model fitted to ranges 12–35 and 0–12 μg/m^3. Top to bottom: models that include lagged meteorology; models that exclude lagged meteorology; various sensitivity analyses.

3 Results

Figure 3 shows estimates and 95% confidence intervals of the $PM_{2.5}$–mortality association under numerous assumptions. The top block of six rows shows the result in different subpopulations for our full model that includes lagged meteorology. The second block shows the same estimates for a model that includes day-of-death meteorology but not lagged meteorology. The third block shows a few sensitivity analyses: including the last day of month (LDOM) and all comparisons days which were earlier omitted; using monitors instead of the RSIG to estimate $PM_{2.5}$; and using 1-day and 5-day averages of lagged meteorology instead of 3-day. For models including lagged meteorology, the regression coefficients were positive and statistically significant in the linear concentration-response function fitted to the full range of $PM_{2.5}$, and for the broken-stick model above 12 $\mu g/m^3$, but not for the broken-stick model below 12 $\mu g/m^3$. When lagged meteorology was omitted, the coefficients were larger across the board and statistically significant both above and below 12 $\mu g/m^3$. The sensitivity analyses showed: including the LDOM has little effect on the results; using monitors instead of RSIG (with correspondingly reduced data coverage) leads to substantially wider confidence intervals, but still with the same relationship between the results with and without lagged meteorology; using 1-day lagged meteorology does not fully account for the confounding effect while 5-day lagged meteorology produces results very similar to those for 3-day lags.

Analyses by race and by region are contained in Figure 2; these analyses were conducted only under the log-linear model but results both excluding and including lagged meteorology are plotted side by side to allow a direct comparison. For the racial comparison, the results show that non-whites are at a higher risk than whites under both models, but with substantially wider confidence intervals for non-whites reflecting the smaller overall population. For the regional analyses, excluding the North-East, the results still show a statistically significant positive effect under the model without lagged meteorology, but this effect disappears under the lagged meteorology model, where none of the estimated coefficients are significantly different from zero. For the specific case of California, the estimated linear coefficient was very close to zero, confirming a result of [32] that was derived using entirely different data and different statistical methodology.

Figure 3 shows plots of the relative risk v. $PM_{2.5}$ curve both with and without lagged meteorology. This curve was computed for the 75+ age group so that the entire analysis could be fitted

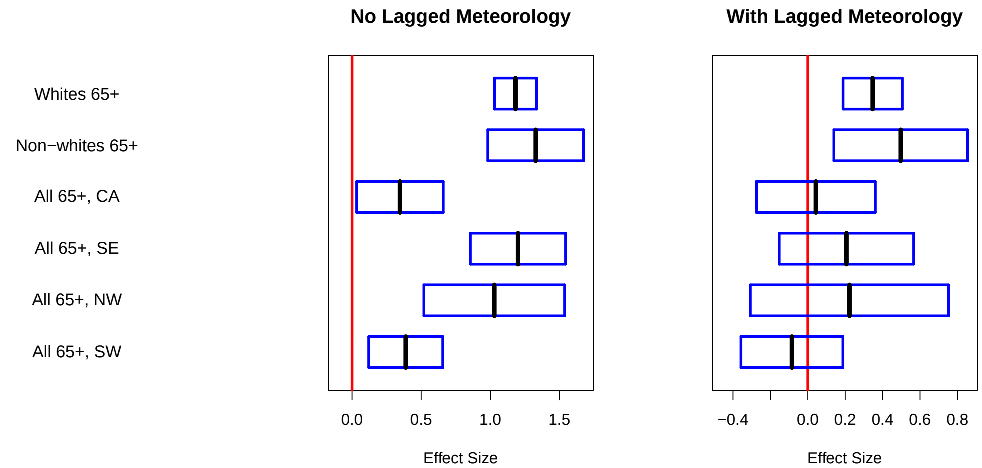

Figure 2: Analyses using the log-linear model by race and by region (other than North-East).

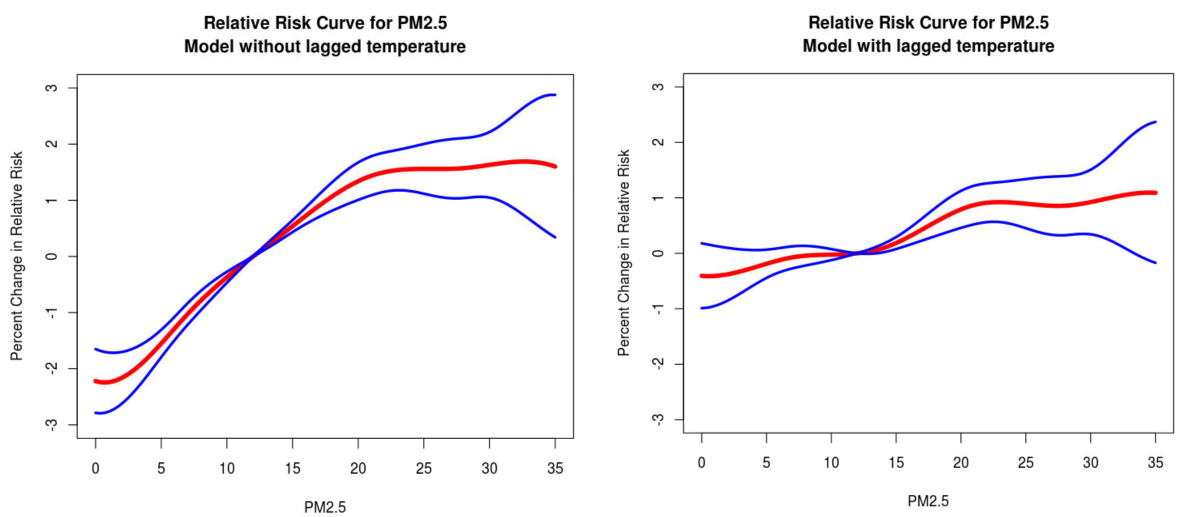

Figure 3: Left curve: Nonlinear relationship between $PM_{2.5}$ and mortality based on analysis without lagged meteorology applied to Medicare enrolees aged 75+. Expressed as percent change in relative risk to a baseline level of 12 $\mu g/m^3$ with pointwise 95% confidence limits. Right: Same curve but including lagged meteorology in the statistical model.

with a single optimization. These plots have been drawn relative to a reference level of 12 μg/m^3, which seems logical because that is the current standard. Relative risks both above and below that standard give some indication of the likely costs or benefits of changing the standard. These curves again demonstrate the importance of taking lagged meteorology into account.

4 Discussion

The results of this paper may be compared with those of an earlier paper [6] that used similar methodology.

In many respects, these results support and validate those of [6]. The results under the No Lags model that are directly comparable with results in [6] show very similar point estimates and confidence intervals. This is despite the fact that the temperature and $PM_{2.5}$ components of the data have been constructed from different sources, and there are differences in the statistical approach as well, including an entirely different computational strategy. The mortality data are similar but not identical, because of slightly different time periods and the end of month anomaly. The present results show a stronger effect in women than in men; increasing effects as age increases; and a stronger association for non-whites than for whites, though with wider confidence intervals for non-whites reflecting differences in population size. All these results are consistent with [6].

Even in the With Lags model, the present results confirm the statistical significance of an overall linear effect in all the national analyses and in the regional analysis for the North-East. However, the results in the 0–12 μg/m^3 range are much attenuated and not statistically significant at the 0.05 level, and the nonlinear concentration-response curve in the right hand half of Figure 3 also implies absence of a statistically significant effect in this range.

There is no clear-cut way of saying which of the two analyses (with or without lagged meteorology) is more appropriate. The most familiar interpretation is that lagged meteorology is acting as a confounder of the $PM_{2.5}$ effect, but this is unclear as there is no obvious mechanism at work here. There would be a direct mechanism if lagged meteorology (along with day of death meteorology) were a reliable predictor of $PM_{2.5}$; but a direct attempt to verify this, using nonlinear splines of both lagged and day of death meteorology as a predictor of $PM_{2.5}$, suggested an R^2 of only 0.14, compared with 0.07 for predicting $PM_{2.5}$ from day of death meteorology alone. Recalling the size of the dataset, even these values are large enough to suggest the possibility of a confounding effect. It is conceivable that the confounding may be the other way round, that $PM_{2.5}$ is acting as a confounder of lagged meteorology, but the statistical significance of lagged meteorology is massive (deviances of several hundred), so any confounding effect by $PM_{2.5}$ could be no more than a very small component of it.

The regional results mimic corresponding results for ozone [30] and are also consistent with the previous result for California in [32]. Possible explanations include different compositions of $PM_{2.5}$ in different parts of the country and also different exposure patterns, e.g. greater use of air conditioning in western and southern states and higher use of public transport in the North-East. It is a matter of speculation whether the different results for California are in any way explained by the fact that California has historically enforced its own air quality standards that are different from the rest of the USA.

The method of this paper does not use "causal inference" techniques in the sense of for example [21, 12, 35], but even in the absence of formal proof of causality, there are strong reasons for believing that the effects are causal. The case-crossover method of analysis, relying on the comparison of date of death with a set of comparison dates for the same individual, practically eliminates any possibility of confounding by individual factors such as physical conditioning, weight, smoking and drug use. There is a possibility of confounding by (a) other meteorological variables besides temperature and dewpoint, and (b) other air contaminants such as certain components of $PM_{2.5}$ having a stronger effect than $PM_{2.5}$ itself. The former possibility has been extensively discussed without reaching clear-cut conclusions; previous analyses such as [31, 29, 32] used specific and relative humidity, atmospheric pressure and separate daily minimum and maximum temperatures. As for the possibility that specific components of $PM_{2.5}$ may be more strongly associated with mortality than overall $PM_{2.5}$, that remains a topic of active research, but at present, only $PM_{2.5}$ itself has a national network.

Numerous caveats remain. These results apply only to short-term mortality; parallel results for long-term mortality were published by [7]. The main part of this analysis covers only 2002-2013; with restrictions due to nonavailability of meteorology and $PM_{2.5}$ data in all zip codes, and further deletions resulting from the last day of month artifact, the analysis dataset comprised less than half the data potentially available from 1999-2018. If all these gaps were filled in with no change in the estimated coefficients, the results could be statistically significant even below 12 μg/m^3.

This paper confirms the statistical significance and likely causality of the overall association between $PM_{2.5}$ and short-term mortality, even under a lagged meteorology model. The possibility that such effects persist to the range of $PM_{2.5}$ below 12 μg/m^3 is by no means ruled out, but further analyses with larger datasets would be needed to resolve this question.

5 Acknowledgements

The staff of the Advanced Biomedical IT Core at Indiana University are thanked for computer services and data storage. The project was managed by Gradient. An earlier version of this analysis was presented at a symposium on Causal Methods in Epidemiological Studies of Particulate Matter and Mortality in Chapel Hill, NC, on October 4, 2018. Comments presented at that symposium significantly impacted the development of the analysis into the present paper. The author is grateful to Dr. Lianne Sheppard for helpful advice about the case-crossover method.

References

[1] J.S. Apte, M. Brauer, Cohen A.J., M. Ezzati, and C.A. Pope. Ambient PM2.5 reduces global and regional life expectancy. *Environmental Science and Technology Letters*, 114 (1):29–33, 2018.

[2] Rob Beelen et al. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. *The Lancet*, 383:785–795, 2014.

[3] M.L. Bell and F. Dominici. Effect modification by community characteristics on the short-term effects of ozone exposure and mortality in 98 US communities. *American Journal of Epidemiology*, 167:986–997, 2008.

[4] M Brauer, JR Brook, T Christidis, Y Chu, DL Crouse, and A. Erickson et al. *Mortality–Air Pollution Associations in Low-Exposure Environments (MAPLE): Phase 1.* Health Effects Institute, Research Report 203, Boston, MA, 2019.

[5] J. Dai, L. Zanobetti, P. Koutrakis, and J. Schwartz. Associations of fine particulate matter species with mortality in the United States: a multicity time series analysis. *Environmental Health Perspectives*, 122 (8):837–842, 2014.

[6] Qian Di, Lingzhen Dai, Yun Wang, Antonella Zanobetti, Christine Choirat, Joel D. Schwartz, and Francesca Dominici. Association of short-term exposure to air pollution with mortality in older adults. *JAMA*, 318 (24):2446–2556, 2017.

[7] Qian Di, Yun Wang, Antonella Zanobetti, Yun Wang, Petros Koutrakis, Christine Choirat, Francesca Dominici, and Joel D. Schwartz. Air pollution and mortality in the Medicare population. *New England Journal of Medicine*, 376 (26):2513–2522, 2017.

[8] F. Dominici, R. Peng, and M. Bell. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA*, 295 (10):1127–1134, 2006.

[9] F. Dominici, J.M. Samet, and S.L. Zeger. Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy. *Journal of the Royal Statistical Society, Series A*, 163(3):263–302, 2000.

[10] F. Dominici, J. Schwartz, Q. Di, D. Braun, C. Choirat, and A. Zanobetti. *Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Phase 1.* Health Effects Institute, Research Report 200, Boston, MA, 2019.

[11] M. Franklin, A. Zeka, and J. Schwartz. Association between PM2.5 and all-cause and specific-cause mortality in 27 US communities. *Journal of Exposure Science and Environmental Epidemiology*, 17 (3):279–287, 2007.

[12] M.A. Hernán and J.M. Robins. *Causal Inference.* Boca Raton: Chapman & Hall/CRC, forthcoming, 2019.

[13] Health Effects Institute. *Revised Analyses of Time-Series Studies of Air Pollution and Health. Special Report.* Health Effects Institute, Boston MA., 2003.

[14] H. Janes, L. Sheppard, and T. Lumley. Case-crossover analyses of air pollution exposure data — referent selection strategies and their implications for bias. *Epidemiology*, 16(6):717–726, 2005.

[15] H. Janes, L. Sheppard, and T. Lumley. Overlap bias in the case-crossover design, with application to air pollution exposures. *Statistics in Medicine*, 24(2):285–300, 2005.

[16] D. Krewski, R.T. Burnett, M.S. Goldberg, K. Hoover, J. Siemiatycki, M. Jerrett, M.h. Abrahamowicz, and W.H. White. Reanalysis of the Harvard Six Cities study and the American Cancer Society study of particulate air pollution and mortality. *Special Report*, Health Effects Institute, Boston, MA, 2000.

[17] D. Krewski, M. Jerrett, R T. Burnett, R. Ma, E. Hughes, Y. Shi, M. C. Turner, C. A. Pope, G. Thurston, E. E. Calle, and M.J. Thun. Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality. *Research Reports of the Health Effects Institute*, 140:5–114, 2009.

[18] D. Levy, T. Lumley, L. Sheppard, J. Kaufman, and H. Checkoway. Referent selection in case-crossover analyses of acute health effects of air pollution. *Epidemiology*, 12(2):186–192, 2001.

[19] M. Maclure. The case-crossover design: a method for studying transient effects on the risk of acute events. *American Journal of Epidemiology*, 133(2):144–153, 1991.

[20] B. Ostro, R. Broadwin, S. Green, W.Y. Feng, and M. Lipsett. Fine particulate air pollution and mortality in nine California counties: results from CALFINE. *Environmental Health Perspectives*, 114 (1):29–33, 2006.

[21] Judea Pearl. *Causality: Models, Reasoning and Inference (second edition)*. Cambridge University Press, 2009.

[22] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in Fortran: The Art of Scientific Computing (Second Edition)*. Cambridge University Press, 1992.

[23] J.M. Samet, F. Dominici, F.C. Curreiro, I. Coursac, and S.L. Zeger. Fine particulate air pollution and mortality in 20 US cities 1987-1994 (with discussion). *New England Journal of Medicine*, 343:1742–1757, 2000.

[24] J.M. Samet, S.L. Zeger, F. Dominici, F. Curriero, I. Coursac, D.W. Dockery, J. Schwartz, and A. Zanobetti. The National Morbidity, Mortality, and Air Pollution Study Part II: Morbidity

and Mortality from Air Pollution in the United States. *Research Report 94-II*, Health Effects Institute, Boston, MA, 2000.

[25] S.E. Sarnat, B.A. Coull, J. Schwartz, D.R. Gold, and H.H. Suh. Factors affecting the association between ambient concentrations and personal exposures to particles and gases. *Environmental Health Perspectives*, 114 (5):649–654, 2006.

[26] J. Schwartz, D. Dockery, and L.M. Neas. Is daily mortality associated specifically with fine particles? *Journal of the Air and Waste Management Association*, 46:927–939, 2000.

[27] J. Schwartz, F. Laden, and A. Zanobetti. The concentration-response relation between PM2.5 and daily deaths. *Environmental Health Perspectives*, 110 (10):1025–1029, 2002.

[28] L. Sheppard. Environmental epidemiology study designs. *Handbook of Environmental and Ecological Statistics, edited by A. Gelfand et al., Chapman and Hall/CRC Press*, Chapter 26:603–616, 2019.

[29] R.L. Smith, J.M. Davis, J. Sacks, P. Speckman, and P. Styer. Regression models for air pollution and daily mortality: analysis of data from Birmingham, Alabama. *Environmetrics*, 11:719–743, 2000.

[30] R.L. Smith, B. Xu, and P. Switzer. Reassessing the relationship between ozone and short-term mortality in U.S. urban communities. *Inhalation Toxicology*, 21 (S2):37–61, 2009.

[31] Patricia Styer, Nancy McMillan, Feng Gao, Jerry Davis, and Jerome Sacks. Effect of outdoor airborne particulate matter on daily death counts. *Environmental Health Perspectives*, 103 (5):490–497, 1995.

[32] S.S. Young, K. Lopiano, and R.L. Smith. Air quality and acute deaths in California, 2000-2012. *Regulatory Toxicology and Pharmacology*, 88:173–184, 2017.

[33] A. Zanobetti and J. Schwartz. The effect of fine and coarse particulate air pollution on mortality. *Environmental Health Perspectives*, 117 (6):898–903, 2009.

[34] Scott L. Zeger, Francesca Dominici, Aidan McDermott, and Jonathan M. Samet. Mortality in the Medicare population and chronic exposure to fine particulate air pollution in urban centers (2000–2005). *Environmental Health Perspectives*, 116(12):1614–1619, 2008.

[35] C.M. Zigler, C. Kim, C. Choirat, J.B. Hansen, Y. Wang, L. Hund, J. Samet, G. King, and F. Dominici. Causal inference methods for estimating long-term health effects of air quality regulations. *Research Report 187*, Health Effects Institute, Boston, MA, 2016.

Dependence of Short-Term Mortality on Fine Particulate Matter in the Population of Elderly Medicare Beneficiaries: Supplementary Material

Richard L. Smith
Department of Statistics and Operations Research,
University of North Carolina, Chapel Hill

November 17, 2021

1 Further Details of Meteorological and $PM_{2.5}$ Data

Meteorology data was obtained from the Global Summary Of the Day (GSOD) database maintained by the National Centers for Environmental Information, a branch of the National Oceanic and Atmospheric Administration (NOAA). GSOD was preferred to the better known Global Historical Climatological Network (GHCN), because GSOD includes dewpoint whereas GHCN does not. Although it is open to discussion whether weather station data is inherently superior to reanalysis data constructions such as NCEP, we note that the NCEP grid resolution is 2.5^o latitude and longitude (roughly 270×210 km. at latitude 40^oN), most points in the US have several weather stations within that distance range, so the weather station data should be more sensitive to spatial variability. The weather variables used for this study were daily mean temperature and dewpoint, chosen so as to match the variables used in [4]. Other variables available in the database include daily maximum and minimum temperature, precipitation and air pressure, and all of these could potentially be included as meteorological covariates as well. [6] found that including daily maximum and minimum temperature separately gave stronger meteorological associations than daily mean temperature, but that possibility is not pursued here. Data were downloaded as gzipped directories (one for each year). Each station is identified by two codes, the USAF code and the WBAN code.

A separate "isd-history" file is available, which lists the same code with other identifiers including latitude-longitude coordinates and a country code. Only stations with a country code of "US" are relevant to the present exercise.

To match the zip code data with weather station and air pollution monitor data, we need latitude-longitude for each zip code and this was obtained from the file `free-zipcode-database-Primary.csv` downloaded from the website `http://federalgovernmentzipcodes.us/`. This dataset lists a total of 42,522 zipcodes with name of city and state, latitude and longitude to the nearest 0.01^o, and some other information. Altogether, 23,796,902 participants out of the original 23,808,244 had zip codes whose geographic coordinates were identifiable in this way, a coverage rate of 99.95% covering 34,553 zipcodes.

For each participant zipcode for which latitude-longitude coordinates were available, the nearest weather station in the GSOD database was identified. Zipcodes for which the distance to the nearest weather station was more than 100 km. were treated as missing — these included data from noncontinental locations such as Guam, for which the distance to the nearest US weather station was in some cases several thousand km. However, such individuals comprise only a tiny fraction of the Medicare database, so their omission should not affect the analysis results in any meaningful way. The choice of 100 km. as the cutoff distance is somewhat arbitrary of course, but [4] used 50 km. for a similar cutoff with $PM_{2.5}$ monitor stations — temperature fields are in general smoother than $PM_{2.5}$ fields, and it seemed reasonable to adopt twice as large a cutoff for the meteorology variables compared with $PM_{2.5}$. The remainder of the meteorology data construction consisted of associating each zipcode with daily temperature and dewpoint data from the nearest weather station, for all days for which these data were available.

This project uses $PM_{2.5}$ data from two sources. Daily data from monitors are available on the EPA website `https://aqs.epa.gov/aqsweb/airdata/download_files.html`, together with ancillary information including latitude-longitude coordinates of the monitors. This was processed in a similar way to the meteorology data, associating each zipcode with the nearest monitor, using a cutoff of 50 km. as previous described. This dataset was limited in two ways: first, each zipcode that is further than 50 km. from the nearest monitor is treated as missing, and second, days where the monitor data are missing are also treated as missing at the associated zipcode. This is a rather significant limitation, as EPA regulations require only that $PM_{2.5}$ be monitored every third day, and the method of case-crossover analysis that we are using for the statistical analysis requires matched $PM_{2.5}$ values at multiples of seven days before and after the day of death. The conflict between these two requirements essentially means that only monitors with daily data are usable, which is a small proportion of the total set of monitors. Therefore, results using $PM_{2.5}$ monitor data will be extremely limited, though they are still useful for validating the results from the fused dataset that is described next.

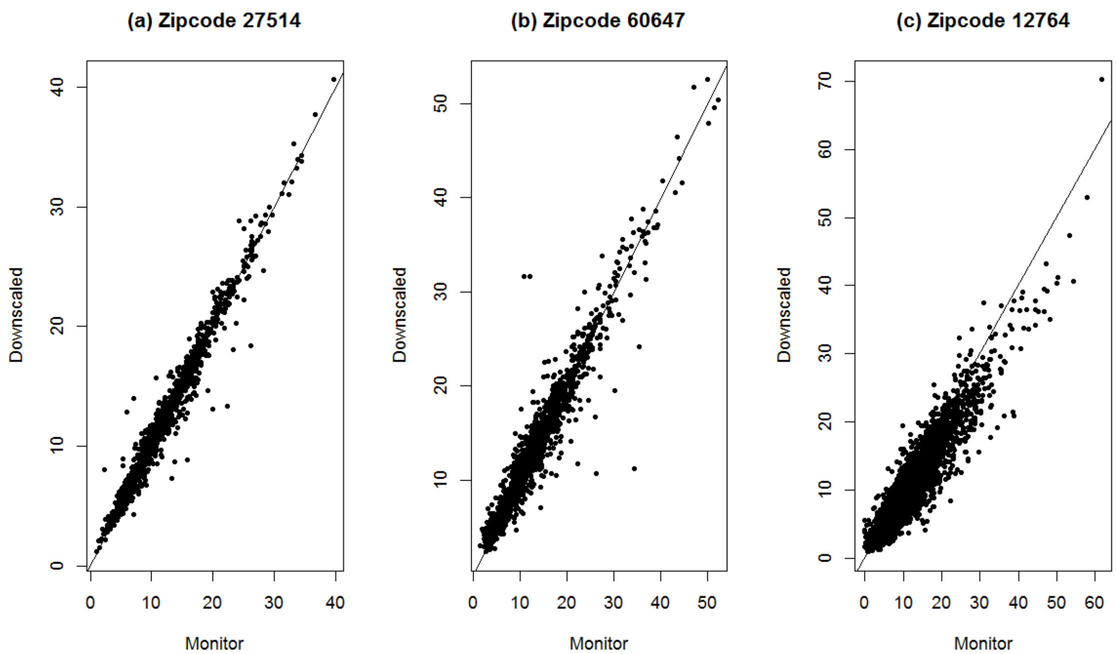

Figure 1: Downscaled versus monitor plots for three zipcodes, 2002–2013. The correlation coefficients for (a)–(c) are respectively 0.98, 0.96 and 0.94.

EPA's Remote Sensing Information Gateway, or RSIG, combines data from monitors and the CMAQ air quality modeling product through a Bayesian approach due to [3, 1, 2]. Daily datasets from 2002 onwards are online at the web address `https://www.epa.gov/hesc/rsig-related-downloadable-data-files`. The data are at the spatial resolution of census tracts, and include both posterior means and posterior standard deviations from the Bayesian algorithm. We shall not use the standard deviations. Files identifying census tracts with zipcodes are available from the website `https://www.huduser.gov/portal/datasets/usps_crosswalk.html`. For example, zipcode 27514 (the author's home zipcode) is identified with 14 census tracts. In this database, the largest number of census tracts in the continental US associated with a single zipcode is 53 (zipcode 60647, Chicago, IL).

As illustrations of the agreement between these two datasets, Figure 1 shows three plots that directly compare the downscaled and monitor data for the zipcodes of 27514 (Chapel Hill, NC), 60647 (Chicago, IL) and 12764 (Narrowsburg, NY), which was picked out because the distance to the nearest monitor is 49.992 km., very close to our 50 km. cutoff. In all cases, the overall agreement is good, though it appears that the downscaled $PM_{2.5}$ is systematically lower than the monitored $PM_{2.5}$ in zipcode 12764, which presumably reflects the greater distance to the monitor for this particular zipcode.

2 Proportions of subjects for which $PM_{2.5}$ on day of death was higher than for any of the three matched days in the case-crossover analysis

Under the case-crossover sampling design, each date of death is matched with three other days in the same 28-day window. If there were no association with $PM_{2.5}$, the proportion for which the day of death has the largest $PM_{2.5}$, among the day of death and the three matched days, should average out to 0.25. This proportion was computed, splitting ties at random. Overall, it comes to 0.252. Based on the sample size of nearly 16 million, this is statistically significant with a very low p-value ($z = 18.5,\ p \approx 10^{-76}$). Even if we restrict to the 4.9 million subjects for whom the day of death and three matched days are all at or below 12 μg/m^3, the observed proportion is 0.251 ($z = 4.92,\ p \approx 10^{-6}$). However there are two caveats about such simplified analyses: the results are much more variable when broken into categories by age, sex, race and region; and, they take no account of possible meteorological confounding. Details are as follows.

For each subject, we calculate the mean $PM_{2.5}$ on the day of death and the day preceding death (the measure of $PM_{2.5}$ used throughout this paper), and then make the corresponding calculation for each of the three matched days. If this value is not available for all four dates, the value is not used. The proportion of subjects for whom the day of death has the largest $PM_{2.5}$ among these four days is calculated. Since $PM_{2.5}$ values are rounded to only one decimal place, there are many ties, but these are resolved by randomly splitting the tie. Thus, if there were no association with $PM_{2.5}$, the proportion of subjects for whom $PM_{2.5}$ on day of death is higher than for any of the three comparison dates should be one quarter. Observed proportions, broken down by the same categories as Table 2 of the main paper, are given in Table 2. Two-sided p-values are calculated assuming binomial sampling. When the calculation is repeated only for days where all four dates have $PM_{2.5}$ values $\leq 12\mu g/m^3$, the results are in Table 2.

Region	Northeast		Southeast		Northwest		Southwest		
Race	White	Other	WhiteM	Other	White	Other	White	Other	Totals
M, 0–64	0.2522^a	0.2541^b	0.2522	0.2504	0.2501	0.2436	0.2526^a	0.251	0.252^c
F, 0–64	0.251	0.2531	0.2505	0.2519	0.2512	0.259	0.2517	0.2526	0.2516^b
M, 65–74	0.2514^a	0.2524^a	0.2515^a	0.2531^a	0.2501	0.2499	0.2511	0.2515	0.2514^c
F, 65–74	0.2516^a	0.254^b	0.2516	0.2522	0.2524	0.2461	0.2514	0.2495	0.2517^c
M, 75–84	0.2526^c	0.2542^c	0.2509	0.2538^b	0.253^c	0.251	0.2512^a	0.2499	0.252^c
F, 75–84	0.2529^c	0.2537^c	0.2518^b	0.2531^a	0.2546^c	0.2571	0.2495	0.2514	0.2521^c
M, 85+	0.2521^c	0.2537^a	0.253^c	0.2527	0.253^b	0.2618^a	0.2504	0.2513	0.2521^c
F, 85+	0.2538^c	0.2534^b	0.2522^c	0.2527^a	0.2523^b	0.2523	0.2505	0.2491	0.2524^c
Totals	0.2527^c	0.2536^c	0.2518^c	0.2526^c	0.2526^c	0.2523	0.2507^b	0.2506	0.252^c

Table 1: The proportion of deaths for which $PM_{2.5}$ (mean of lag 0 and lag 1) on the day of death was higher than that on any of the three comparison days, splitting ties at random. Bottom right hand entry (0.252) is for all categories combined; the rest of the table makes the same calculations for various subcategorizations of the data. Superscripts indicate level of statistical significance, based on a two-sided significance level of 0.05 (a), 0.01 (b) or 0.001 (c).

3 The "Last Day Of Month" Issue

Initial processing of the data identified an apparent anomaly, that the number of deaths on the last day of each month is substantially higher than on the remaining days of the month. This is illustrated in Figure 2 (plotted for 2002–2013, as these are the years for which the downscaled $PM_{2.5}$ data are available). The effect declines over time and disappears entirely during the final 1.5

Region	Northeast		Southeast		Northwest		Southwest		
Race	White	Other	White	Other	White	Other	White	Other	Totals
M, 0–64	0.2489	0.2549	0.2496	0.2467	0.2513	0.2433	0.2522	0.2495	0.2502
F, 0–64	0.2484	0.2447	0.251	0.2547	0.2522	0.2588	0.252	0.2527	0.2512
M, 65–74	0.2496	0.2452	0.2517	0.2508	0.2515	0.246	0.2534^b	0.2529	0.2514^a
F, 65–74	0.2503	0.2458	0.2501	0.2486	0.2521	0.2419	0.2516	0.2528	0.2507
M, 75–84	0.25	0.2488	0.2488	0.254	0.2552^c	0.2539	0.2506	0.2502	0.2509
F, 75–84	0.2507	0.2535	0.2522^a	0.2524	0.2546^c	0.2587	0.2486	0.2543^a	0.2515^b
M, 85+	0.2494	0.2443	0.2526^a	0.2491	0.2537^b	0.2601	0.2506	0.2494	0.2511^a
F, 85+	0.2494	0.2487	0.2509	0.25	0.253^b	0.2553	0.2506	0.248	0.2506
Totals	0.2497	0.2486	0.251^a	0.2508	0.2534^c	0.2523	0.2508^a	0.2511	0.251^c

Table 2: Same as Table 2, but restricted to deaths for which the $PM_{2.5}$ level on the day of death and on all three comparisons days was less than 12 $\mu g/m^3$.

years of data. A reviewer has pointed out that this is a known feature of Medicare data: deaths without a validated date of death are assigned to the last (or sometimes first) day of the month, and there is a flag variable in the Medpar file that identifies valid dates of death. This flag was either missing or overlooked at the time of compiling the data for the present study. As noted in the main paper, the anomaly was dealt with in our analysis by omitting the last day of month, and all days paired with a last day in the case-crossover analysis, thereby eliminating the bias.

4 Further details of analytic method

For the ith individual, we have covariates x_{ijk} representing covariates $j = 1, ..., 5$ (four weather variables and one $PM_{2.5}$, which may be either downscaled or monitored — we shall not have occasion to combine the two) and comparison days $k = 1, ..., 4$, where $k = 1$ represents day of death and $k = 2, 3, 4$ are the three comparison days (days that are either 7, 14 or 21 days before or after date of death within the same 28-day referent window). In subsequent analysis we will replace the individual x_{ijk} values by expansions using B-splines to represent nonlinear effects. Assuming for the moment that these are given, we may assume a vector of covariates $\mathbf{x}_{i,k}$ for the ith individual

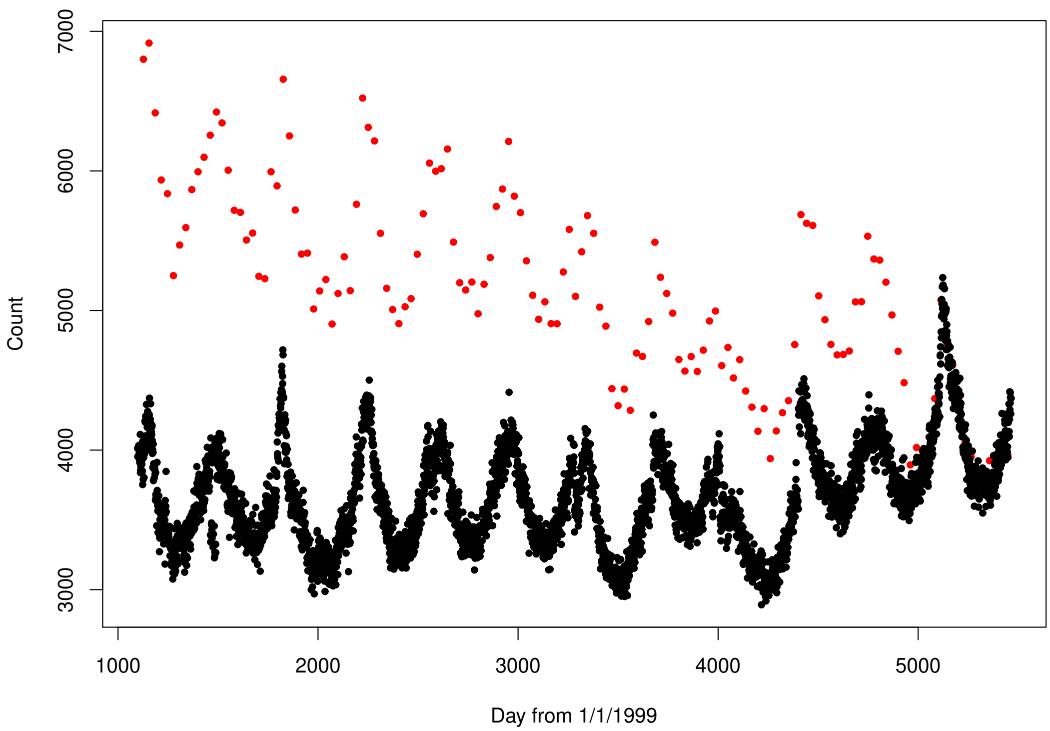

Figure 2: Number of deaths per day plotted against day since 1/1/1999. Last day of month is plotted in red.

on the kth comparison date. The likelihood function [4, 5] is then

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} \frac{\exp(\mathbf{x}_{i,1}^T \boldsymbol{\beta})}{\sum_{k=1}^{4} \exp(\mathbf{x}_{i,k}^T \boldsymbol{\beta})}. \quad (1)$$

Here, N is the number of individuals and $\boldsymbol{\beta}$ represents a vector of model parameters, whicn include both meteorological and $PM_{2.5}$ effects. The function (1) is maximized using standard optimization algorithms.

For the regional analyses mentioned in Section 2 of the main paper, the following definitions are used:

1. US — the whole country
2. CA — California only
3. NE — CT,DE,IL,IN,ME,MD,MA,MI,NH,NJ,NY,OH,PA,RI,VT,WI
4. SE — AL,AR,DC,FL,GA,KY,LA,MS,NC,SC,TN,VA,WV
5. NW — CO,IA,ID,MN,MT,ND,NE,OR,SD,UT,WA,WY
6. SW — AZ,CA,KS,MO,NV,NM,OK,TX

References

[1] V. Berrocal, A.E. Gelfand, and D.M. Holland. A bivariate space-time downscaler under space and time misalignment. *Annals of Applied Statistics*, 4:1942–1975, 2010.

[2] V. Berrocal, A.E. Gelfand, and D.M. Holland. A spatiotemporal downscaler for output from numerical models. *J. of Agricultural, Biological,and Environmental Statistics*, 15:176–197, 2010.

[3] V. Berrocal, A.E. Gelfand, and D.M. Holland. Space-time fusion under error in computer model output: an application to modeling air quality. *Biometrics*, 68:837–848, 2012.

[4] Qian Di, Lingzhen Dai, Yun Wang, Antonella Zanobetti, Christine Choirat, Joel D. Schwartz, and Francesca Dominici. Association of short-term exposure to air pollution with mortality in older adults. *JAMA*, 318 (24):2446–2556, 2017.

[5] L. Sheppard. Environmental epidemiology study designs. *Handbook of Environmental and Ecological Statistics, edited by A. Gelfand et al., Chapman and Hall/CRC Press*, Chapter 26:603–616, 2019.

[6] S.S. Young, K. Lopiano, and R.L. Smith. Air quality and acute deaths in California, 2000-2012. *Regulatory Toxicology and Pharmacology*, 88:173–184, 2017.