NBER WORKING PAPER SERIES

BEYOND MEAN EXPOSURE: HOW EXPOSURE VARIANCE SHAPES AIR POLLUTION MORTALITY RISK

Bissan Ghaddar
Kimberly Singer Babiarz
Hongbin Li
Lingsheng Meng
Lynn Hildemann
Han Hong
Alvise Scarabosio
Meili Wang
Aprajit Mahajan
Peng Yin
Maigeng Zhou
Grant Miller

Working Paper 35428
http://www.nber.org/papers/w35428

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2026

This research was funded by the Stanford Woods Institute for the Environment through the Environmental Venture Projects program. Bissan Ghaddar's research is supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant RGPIN-2025-04585 and by the John Thompson Chair Fellowship. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Beyond Mean Exposure: How Exposure Variance Shapes Air Pollution Mortality Risk
Bissan Ghaddar, Kimberly Singer Babiarz, Hongbin Li, Lingsheng Meng, Lynn Hildemann, Han Hong, Alvise Scarabosio, Meili Wang, Aprajit Mahajan, Peng Yin, Maigeng Zhou, and Grant Miller
NBER Working Paper No. 35428
July 2026
JEL No. O10, O13

ABSTRACT

Ambient air pollution contributes substantially to premature mortality globally, and the effectiveness of policies addressing it depends critically on the shape of the dose-response relationship, which remains poorly understood. Linking hourly air pollution data with daily vital statistics from 251 Chinese cities over four years, we show this relationship depends not only on the mean, but also on the variance of within-period exposure. At a 7-day mean of 150 μg/m³, predicted mortality varies more across observed hourly PM2.5 distributions than across mean exposure differences of 100 μg/m³. The sign of this variance effect also depends on the mean: at low mean concentrations (a convex region), higher variance is associated with modestly higher mortality. At higher concentrations (a concave region), greater variance is associated with lower mortality – implying constant exposure is more harmful than fluctuating exposure for a given mean. These findings reveal important relationships masked using temporally aggregated air pollution data.

Bissan Ghaddar
Western University
Ivey Business School
bissan.ghaddar@uwaterloo.ca

Kimberly Singer Babiarz
Stanford University
babiarz@stanford.edu

Hongbin Li
Stanford University
Stanford Institute for Economic Policy Research (SIEPR)
Department of Economics
lihongbinsem@gmail.com

Lingsheng Meng
Stanford University
lmeng@stanford.edu

Lynn Hildemann
Stanford University
Civil and Environmental Engineering
hildemann@stanford.edu

Han Hong
Stanford University
doubleh@stanford.edu

Alvise Scarabosio
Stanford University
alvise@stanford.edu

Meili Wang
Yale University
meili.wang@yale.edu

Aprajit Mahajan
University of California, Berkeley
and NBER
aprajit@gmail.com

Peng Yin
Chinese Center for Disease Control
and Prevention
yinpeng@ncncd.chinacdc.cn

Maigeng Zhou
Chinese Center for Disease Control
and Prevention
zhoumaigeng@ncncd.chinacdc.cn

Grant Miller
Stanford University
School of Medicine
Department of Health Policy
and NBER ngmiller@stanford.edu

1 Introduction

Ambient air pollution contributes substantially to premature mortality around the world. $PM_{2.5}$ pollution alone is estimated to reduce life expectancy at birth by roughly 1 year globally, by 1.2-1.9 years in heavily polluted regions of Asia and Africa, and by nearly 5 years in the most heavily polluted areas of the world (1; 2). Research on China (the context of our analysis) suggests that a $10\,\mu g/m^3$ increase in ambient PM_{10} concentration reduces life expectancy by 0.64 years (3). Measured directly in deaths, $PM_{2.5}$ pollution is associated with 3-4.8 million global deaths each year including infant deaths, and is ranked as the fifth most important mortality risk factor (4; 5; 6; 7).

The effects of policy efforts to reduce ambient air pollution—as well as the resources that should be devoted to them – depend critically on the shape of the dose-response relationship between air pollution exposure and mortality (8; 9; 10). Prior work has examined this relationship by relating mortality to *mean* $PM_{2.5}$ concentrations, typically at the annual or daily level (11; 12; 13; 14). This approach implicitly treats the mean as a "sufficient statistic" for the exposure distribution's effect on health. In other words, two periods with the same mean $PM_{2.5}$ exposure are taken to imply the same mortality risk – regardless of how exposure varies within the period. This reliance on mean exposure is partly due to data availability. Many prominent studies rely on annual mean concentration data (6; 15), and although more recent work uses daily means (16; 17) intra-day fluctuations are masked. Research using more disaggregated hourly data has generally been confined to small geographic samples (18; 19).[1] However, whether or not the mean is in fact a sufficient summary statistic for studying the relationship between air pollution exposure and mortality, and how within-period variation in exposure independently affects health outcomes, remains largely unexplored at scale.

Within this mean exposure framework, the shape of the dose-response curve has received considerable attention. Considerable evidence suggests that it appears concave at high concentrations. For example, long-term studies in China (15) and studies of acute smoke events in the United States

[1] There is only one population-level study of which we are aware that examines within-period variation directly – Lin et al. (20), which studies the standard deviation of daily $PM_{2.5}$ exposure and respiratory mortality in Hong Kong.

(17; 21) document health effects that plateau at high levels.[2] Concavity in this range implies that the incremental health benefits of pollution reduction in highly polluted regions may be small, and that large reductions are needed to meaningfully improve health (4; 23; 24; 25; 9; 26). Evidence of convexity at lower concentrations – under which even small reductions yield substantial benefits – is by contrast limited (27; 28). However, interpretations of either shape of the dose-response curve generally assume that a single mean-concentration curve is informative, regardless of within-period variation in exposure.

In this paper, we provide direct evidence on whether or not mean $PM_{2.5}$ exposure is in fact a sufficient summary statistic of exposure for assessing mortality risk. We find that it is not: at a given mean, predicted mortality varies substantially across the hourly distribution of exposure, and the sign of this variance effect depends on the level of mean exposure. We show this using hourly ambient air pollution data (including $PM_{2.5}$ concentrations) linked with daily vital statistics and meteorological data from 251 Chinese cities over a four-year period (2013–2016). To estimate the relationship between $PM_{2.5}$ exposure and mortality without imposing a parametric functional form, we use a data-driven machine learning approach (XGBoost) to identify concentration cutpoints, and we then estimate the dose-response relationship with regressions using these cutpoints to measure the share of exposure hours in each concentration interval. Finally, with the resulting estimates, we trace out predicted mortality across a wide range of exposure mean and variance combinations. We find that the mean-concentration curve appears convex at low mean $PM_{2.5}$ exposure and concave at high mean $PM_{2.5}$ exposure, consistent with prior work. However, the curve across the highest-variance quartile of observations at each mean is nearly linear, while the same curve across the lowest-variance quartiles exhibits a strong convex-then-concave shape. An important implication of these results is that at higher mean exposure, persistent (low-variance) exposure is more harmful than fluctuating (high-variance) exposure with the same mean – a direct consequence of the concavity at high concentrations.

[2] Proposed mechanisms for this flattening include both physiological saturation of the body's response and protective behavioral responses by individuals to limit exposure during extreme pollution events (22).

2 Results

2.1 Partitioning Exposure and Relationships with Mortality Risk

We first partition the spectrum of $PM_{2.5}$ concentrations into empirical-determined concentration intervals with distinct mortality impacts. To do so, we first use a data-driven approach based on XGBoost, a machine learning algorithm which ranks features according to their contributions to improving predictive accuracy. Our model features are the share of hours within a reference period (7 days or 28 days, loosely corresponding to periods of one week or one month) for which $PM_{2.5}$ measurements were at or above a given level, along with daily average temperature, humidity, and air pressure. Our cross-fitting procedure (Appendix Section A1.2.1) applies XGBoost separately to each of two city-level folds (Fold A and Fold B, a random split of cities), producing a fold-specific set of thresholds for each window. For the 7-day window, Fold A's thresholds are 110, 130, and 150 $\mu g/m^3$ and Fold B's are 70, 100, 130, and 150 $\mu g/m^3$; for the 28-day window, Fold A's thresholds are 80, 100, 130, and 150 $\mu g/m^3$ and Fold B's are 80, 130, and 150 $\mu g/m^3$ (all values fall within the standard air quality index (AQI) unhealthy-for-sensitive-groups to unhealthy band). Using each fold's thresholds, we measure the share of hours in the resulting concentration intervals over the preceding 7 or 28 days, and then estimate the regression model shown in Equation (A2) (controlling for year and city fixed effects, daily mean temperature, humidity, and air pressure) on the held-out fold.

Our cross-fitted OLS estimates from Equation (A2) yield marginal mortality effects for each $PM_{2.5}$ concentration interval for 7- and 28-day preceding windows. Across all four fold-by-window regressions, the estimated coefficient on every time-share interval is positive and statistically significant, with incremental effects per percentage-point of time in an interval broadly similar across bins (approximately 0.09 – 0.15 deaths per million per percentage-point). Using these estimates to then predict dose-response curves (Figures 1 and 2), these estimates imply that at a mean $PM_{2.5}$ exposure of 130 $\mu g/m^3$, excess mortality relative to our low-pollution reference ($PM_{2.5}$ = 5 $\mu g/m^3$ with constant exposure) is approximately 0.75 deaths per million per day for the 7-day window

(bootstrap 95% CI: [0.57, 1.07]) and approximately 1.18 deaths per million per day for the 28-day window (bootstrap 95% CI: [0.71, 1.61]). Excess mortality is monotonically increasing in mean $PM_{2.5}$ throughout the range of our data.

Appendix Section A2 shows that our results are robust to different choices of $PM_{2.5}$ category thresholds and model specifications.

2.2 Variability in Dose-Response Curves between Pollution Exposure and Mortality

We then use these estimates to conduct two simulation exercises, tracing out the entire dose-response curves for $PM_{2.5}$ exposure and mortality implied by Equation (A2). To address concerns about overfitting (using the same data for both threshold selection and estimation), we employ a cross-fitting procedure. Specifically, we divide cities into two folds, training our XGBoost model on one fold to identify thresholds, using the other fold for OLS estimation with intervals defined by those thresholds – and then vice versa (see Appendix Section A1.2.1 for details). Figure 1 shows that the resulting dose-response curves are precisely estimated, with tight bootstrapped 95% confidence intervals.

We conduct these exercises for both 7- and 28-day periods (separately), predicting daily mortality using using two types of data: (1) values of $PM_{2.5}$ concentrations actually observed in our data, and (2) simulated values of $PM_{2.5}$ concentrations (to explore mean-variance combinations not observed in our data — see Appendix Section A1.2.4 for more details).

Figure 2A shows predictions for 7-day exposure using observed city-hour observations in our dataset. The red curve depicts mean implied mortality for each mean $PM_{2.5}$ concentration, and the shaded gray region reflects the full range of implied mortality values for all distributions of hourly exposure at each mean $PM_{2.5}$ concentration.

Focusing first on the mean mortality curve (in red) using all observations in our data, Figure 2A shows that the dose-response curve appears largely convex up to about 100 $\mu g/m^3$, and then becomes concave at higher concentrations. This curve implies that on one hand, even small reduc-

tions in ambient $PM_{2.5}$ pollution below 100 $\mu g/m^3$ can be associated with quantitatively important reductions in mortality. On the other hand, at higher concentration levels, a reduction in $PM_{2.5}$ pollution is associated with successively smaller mortality benefits . In the extreme, at concentrations above 250 $\mu g/m^3$, reductions in ambient $PM_{2.5}$ pollution have essentially no mortality benefit unless they are sufficiently large to reduce concentrations to substantially lower levels. Note, however, that observations become increasingly sparse at concentrations exceeding 250 $\mu g/m^3$, making estimates in this range considerably more uncertain (see Appendix Figure A1).

Turning to the shaded gray regions in Figure 2A, a striking result is how much predicted mortality rates vary at each given level of mean $PM_{2.5}$ exposure – depending on the underlying distribution of hourly exposure. For example, excess mortality implied by a mean exposure of 150 $\mu g/m^3$ varies from approximately 0.36 to 1.27 deaths per million per day (a within-mean difference of 0.91 deaths per million). This variation in mortality, all with a mean value of 150 $\mu g/m^3$, is larger than the difference in mean excess mortality between 70 $\mu g/m^3$ (0.24 deaths per million) and 170 $\mu g/m^3$ (0.93 deaths per million), a difference of 0.69 deaths per million across a 100 $\mu g/m^3$. This comparison highlights the importance of temporal distributions of exposure, beyond mean exposure values.

Because our data is sparse beyond 200 $\mu g/m^3$, Figure 2B then shows the same curve generated using simulated city-hour observations. In general, the results are similar to Figure 2A, showing substantial variation in mortality, depending on the underlying hourly distributions of exposure, at each level of mean exposure. It also shows that at higher mean $PM_{2.5}$ concentrations, there is less variation in predicted mortality due to the underlying hourly distribution of exposure. This narrowing of the gray region at higher mean concentrations is consistent with a concave dose-response curve at high concentrations, but it is also mechanically constrained by the top exposure category (and associated estimate) in our OLS specification. However, Appendix Section A2 and Appendix Figure A4 show that the same pattern of results persists with higher truncation points (we re-estimate the model using higher top exposure cutoffs (200 $\mu g/m^3$ and 250 $\mu g/m^3$) and this yields similar dose-response shapes)—the convex-to-concave transition and dependence of predicted

mortality on the underlying hourly exposure distribution remain qualitatively unchanged.

Panels A and B in Figure 3 then show these relationship using ridgeline plots depicting the conditional distributions of mortality at selected levels of mean $PM_{2.5}$ distribution. At each y-level, the height of the ridge represents the density of mortality predictions across observations at that level of mean $PM_{2.5}$ exposure. The variation in mortality at each mean value of $PM_{2.5}$ exposure is again substantial.

Panels C and D of Figure 2 and 3 show the same ridgeline for 28-day exposure. In general, they are similar to the 7-day exposure curves, with modestly higher mortality levels. In all cases, the shape of the dose-response curves and the amount of variation in mortality at each level of mean $PM_{2.5}$ exposure are similar as well.

2.3 Is Constant or Variable Ambient Air Pollution Exposure More Harmful?

Ultimately, the importance of high-frequency air pollution measurement depends on how much it matters for population-level mortality. Figures 4A and 4B (7-day exposure) and Figures 4C and 4D (28-day exposure) use actual and simulated observations to show how the dose-response relationship between mean $PM_{2.5}$ exposure and predicted mortality risk depends on the variance of exposure over time. For these plots, we summarize temporal variability using the variance of daily mean $PM_{2.5}$ within the 7- (or 28-) day window (i.e., the mean squared deviation of each day's 24-hour average from the window mean), and we generate predicted mortality using Equation A3 together with each underlying distribution of hourly exposure. Specifically, they show separate dose-response curves for each quartile of $PM_{2.5}$ exposure variance across all observed mean levels of exposure. Strikingly, the predicted dose-response curve for the top variance quartile (Q4) appears nearly linear. By contrast, the predicted dose-response curve for the bottom variance quartile (Q1) appears highly non-linear, with a pronounced convex relationship at low mean levels, and a similarly pronounced concave relationship at higher mean levels of $PM_{2.5}$ exposure.

Examining what patterns of hourly exposure are most harmful, the bootstrap confidence intervals on the Q4–Q1 gap (Figure 5) show a sign-flipping pattern. At low mean $PM_{2.5}$ (near 50 $\mu g/m^3$)

the gap is small, but positive with a 95% CI excluding zero, implying modestly higher mortality in the high-variance quartile (Q4). Alternatively, at higher concentrations (roughly 150 $\mu g/m^3$ and above), the gap is large and negative with a confidence interval excluding zero, indicating that constant exposure (Q1) is considerably more harmful than fluctuating exposure (Q4) (with some convergence across quartiles above approximately 200 $\mu g/m^3$, though the gap becomes noisier and less precisely estimated). Overall, these figures suggest that for a given mean exposure at higher $PM_{2.5}$ concentrations, mortality risk is substantially greater with constant exposure than with fluctuating exposure.

This pattern of results can be interpreted as a natural consequence of the dose-response shape shown in Figure 2, with convexity up to approximately 100 $\mu g/m^3$ and then concavity at higher levels. At low mean $PM_{2.5}$ levels over which the curve is convex, additional temporal variability (conditional on a mean level of exposure) shifts some hours to higher concentrations and some hours to lower concentrations. The extra harm from additional hours of higher pollution exposure outweighs the benefit of additional hours of lower pollution exposure, resulting in slightly higher mortality. Alternatively, at high mean levels of $PM_{2.5}$ exposure over which the curve is concave, additional variability moves some hours down from very high concentrations on the plateau to lower levels, where exposure reductions matter more, while pushing other hours higher leads to relatively little additional harm–resulting in lower mortality.[3]

3 Conclusion

Ambient air pollution causes substantial harm to population health around the world, and the value of efforts to reduce air pollution depend critically on the underlying shape of the dose-response relationship between pollution exposure and mortality risk. Environmental policies have long treated mean $PM_{2.5}$ as a sufficient statistic of exposure for mortality risk. In this paper, we show that this approach is incomplete: at a given mean, the within-period distribution of hourly

[3] At very high mean $PM_{2.5}$ concentrations, where most hours are on the concave section of the dose-response curve, reallocating hours of pollution exposure has almost no effect, resulting in convergence of predicted mortality across variance quartiles.

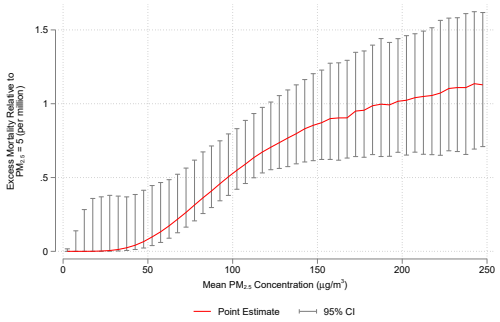
exposure substantially shifts predicted mortality, and the sign of this variance effect depends on the level of mean exposure. Failure to measure and account for high-frequency variation in exposure is therefore a constraint to better understanding the health harms of ambient air pollution and the health benefits of policies to reduce them.

Three findings underlie this conclusion. First, mean $PM_{2.5}$ is not a sufficient statistic of exposure for mortality risk: at a given mean, predicted mortality varies substantially across the hourly distributions consistent with that mean. Second, the predicted dose-response curve for the highest-variance quartile of observations is nearly linear, while the curve for the lowest-variance quartile is highly non-linear. Third, at higher mean exposures – where consistent with prior work (15; 17), the aggregate curve appears concave – persistent (low-variance) exposure is, on average, more harmful than fluctuating (high-variance) exposure, a direct consequence of concavity. Together these findings complement a growing literature identifying within-period exposure variability as an important independent predictor of mortality (20).

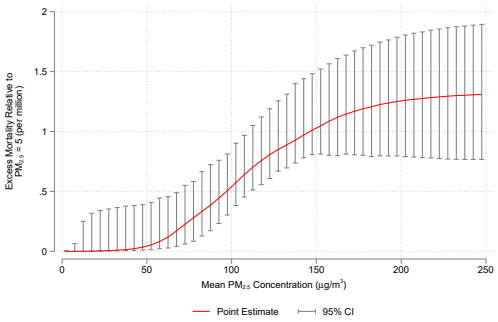
Our paper also has several important limitations. First, our estimates describe associations between $PM_{2.5}$ exposure and mortality rather than causal effects. Although we control for meteorological conditions, city fixed effects, and year fixed effects, unmeasured confounders do not allow us to make causal claims. Second, our estimated dose-response curves are reduced-form relationships that capture both the direct physiological effects of pollution exposure and any behavioral responses to varying pollution levels. For example, individuals may engage in protective behaviors during high pollution episodes (e.g., reducing outdoor activities, wearing masks, or using indoor air filtration) that could attenuate health impacts (22; 21). Our analysis cannot empirically disentangle clinical toxicity at different exposure intensities from behavioral responses that vary with pollution levels. The extent to which such behavioral responses explain the observed patterns in our dose-response curves, particularly the concavity at very high concentrations, is an important question for future research.

Finally, our results have direct implications for air-quality policy. Most national and international ambient air quality standards focus on 24-hour mean concentrations, implicitly treating constant and

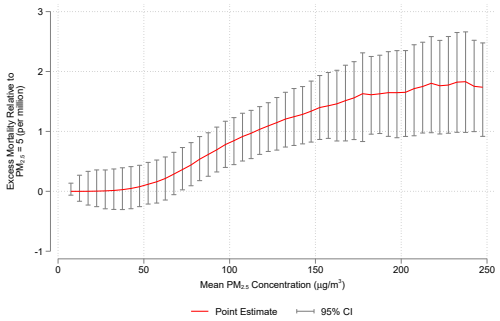
fluctuating exposure at a given mean as equivalent. Our results suggest that this equivalence may not hold: in the highly-polluted regions where most regulatory attention is concentrated, persistent exposure appears, on average, to be more harmful than fluctuating exposure at a given mean. More broadly, for many average values of air pollution commonly observed around the world, the mortality implications of mean exposure may depend almost as much on the within-period variance of exposure as they do on the mean itself. This represents an important area for future research.



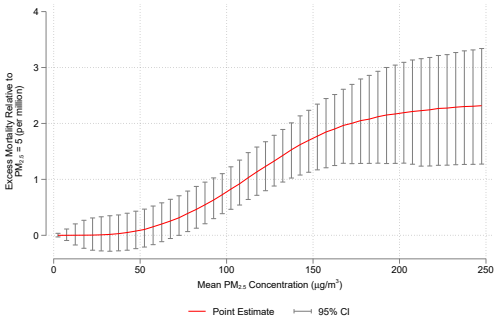
(A) 7-Day Window, Actual Data



(B) 7-Day Window, Simulated Data



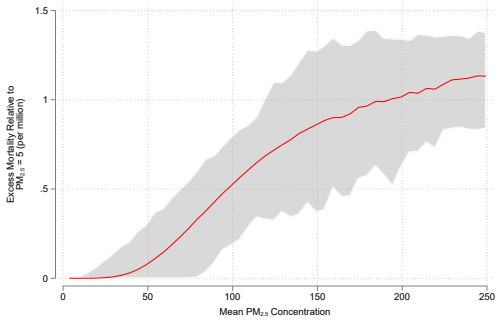
(C) 28-Day Window, Actual Data



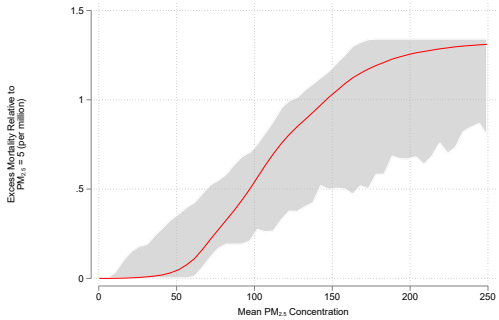
(D) 28-Day Window, Simulated Data

Figure 1: Dose-Response Curve with Bootstrap 95% Confidence Intervals

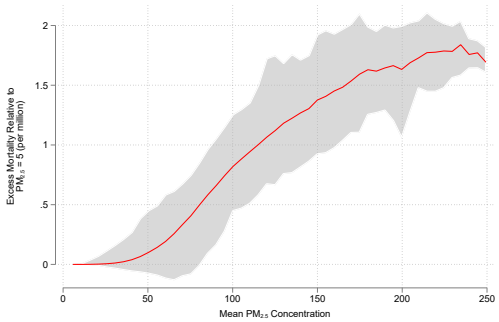
Notes: This figure plots the dose-response relationship between mean $PM_{2.5}$ exposure and predicted excess mortality, with pointwise 95% bootstrap confidence intervals. The solid red line shows the mean predicted excess mortality from the original cross-fitted estimates; the grey whiskers show the 2.5th and 97.5th percentiles across 250 bootstrap iterations. Each iteration resamples cities with replacement within their cross-fitting fold and re-runs the full XGBoost threshold identification and OLS regression pipeline (see Appendix Section A1.2.5). The x-axis represents the mean $PM_{2.5}$ concentration ($\mu g/m^3$) over the preceding 7 days (Panels A, B) or 28 days (Panels C, D). The y-axis is excess predicted daily mortality per 1 million population relative to a reference exposure of $PM_{2.5}$ = 5 $\mu g/m^3$ with constant exposure; at this reference level the curve is zero by construction. Panels A and C use the actual distribution of hourly $PM_{2.5}$ concentrations observed in our sample; Panels B and D use simulated concentrations. Plots are truncated at 250 $\mu g/m^3$.



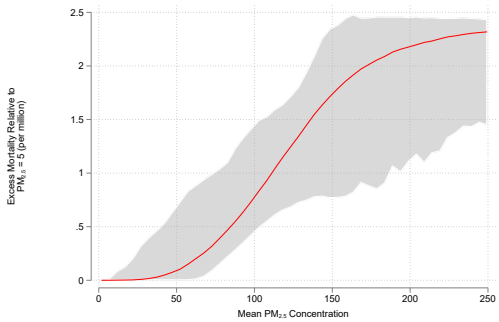
(A) Predicted Mortality: 7-day Actual Exposure



(B) Predicted Mortality: 7-day Simulated Exposure



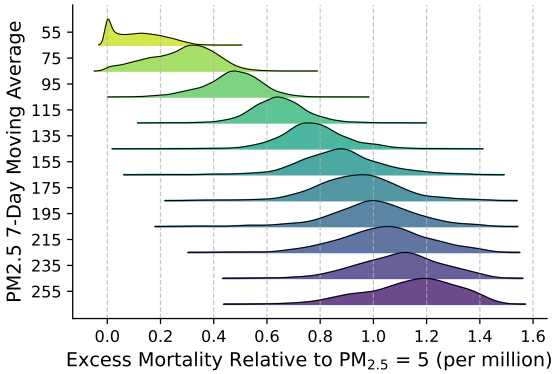
(C) Predicted Mortality: 28-day Actual Exposure



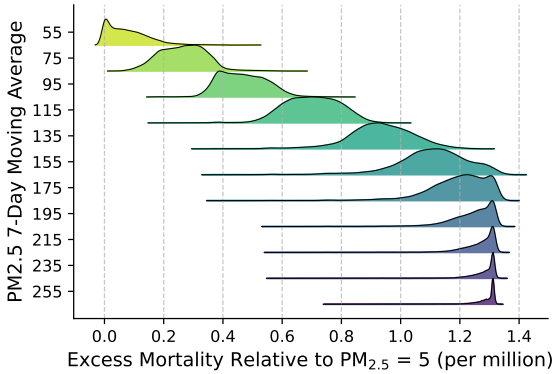
(D) Predicted Mortality: 28-day Simulated Exposure

Figure 2: Predicted Mortality from $PM_{2.5}$ Exposure

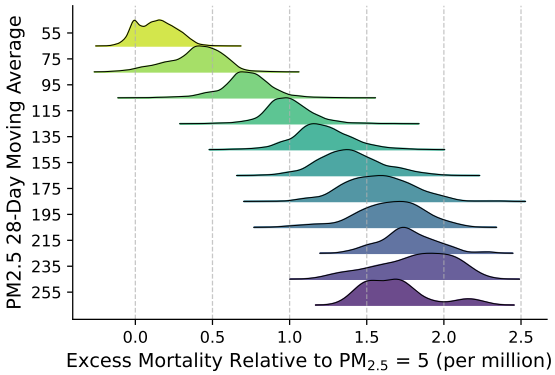
Notes: This figure plots the dose-response relationship between mean $PM_{2.5}$ exposure and predicted mortality. The solid red line shows the mean predicted mortality for each level of average $PM_{2.5}$ exposure. The shaded gray region represents the full range (minimum to maximum) of predicted mortality values arising from all underlying hourly exposure distributions that result in the same mean $PM_{2.5}$ exposure. The x-axis represents the mean $PM_{2.5}$ concentration ($\mu g/m^3$) over the preceding 7 days (Panels A, B) or 28 days (Panels C, D). The y-axis is excess predicted daily mortality per 1 million population relative to a reference exposure of $PM_{2.5}$ = 5 $\mu g/m^3$ with constant exposure. Predictions are generated using the cross-fitted regression coefficients from Equation (A2), applied to the specific distribution of hourly pollution for each city-day observation. Weather variables are held constant at their sample means. Expressing predictions as excess mortality relative to the low-pollution reference isolates the dose-response component from city and year fixed effects. Panels A and C use the actual distribution of hourly $PM_{2.5}$ concentrations observed in our sample. Panels B and D use simulated concentrations. Plots are truncated at a mean exposure of 250 $\mu g/m^3$ due to data sparsity.



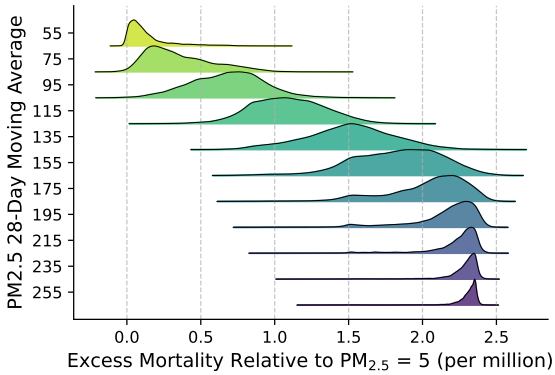
(A) 7-Day Window, Actual Data



(B) 7-Day Window, Simulated Data



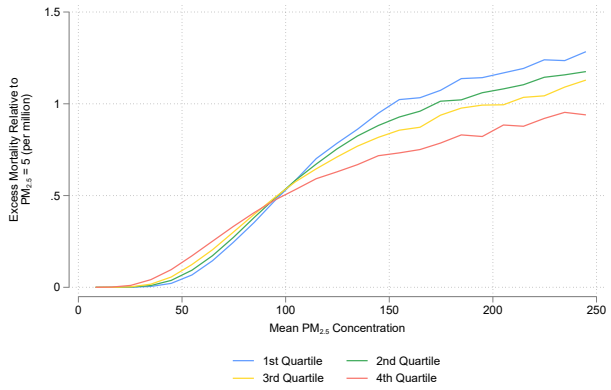
(C) 28-Day Window, Actual Data



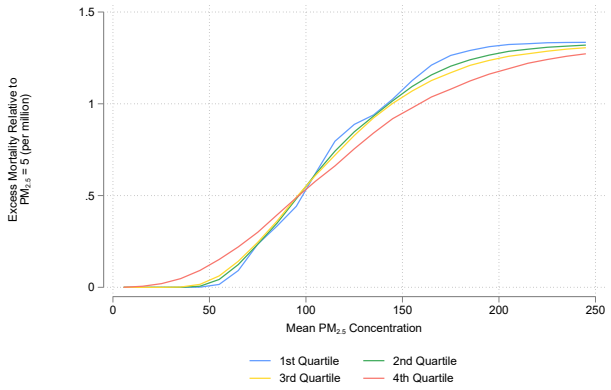
(D) 28-Day Window, Simulated Data

Figure 3: Dose-response Curves Ridgeline Plot for $PM_{2.5}$ Exposure

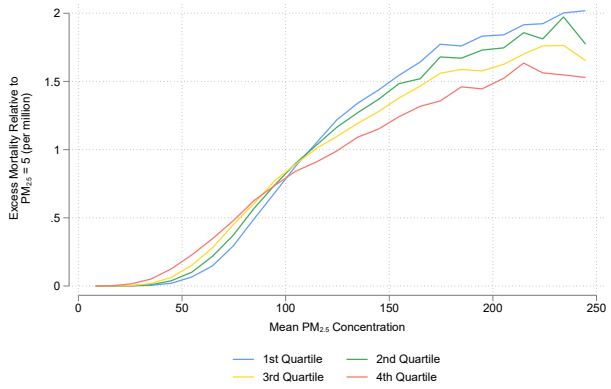
Notes: This figure visualizes the distribution of predicted mortality rates at selected levels of mean $PM_{2.5}$ exposure, illustrating the variability represented by the shaded regions in Figure 2. The shape of the 'ridge' shows the probability density of predicted mortality rates (x-axis) for all underlying hourly exposure patterns that result in a bin centered around the specific mean. Predictions are generated using the cross-fitted regression coefficients from Equation (A2), applied to the specific distribution of hourly pollution for each city-day observation. Weather variables are held constant at their sample means. Predictions are expressed as excess mortality relative to $PM_{2.5} = 5$ $\mu g/m^3$ with constant exposure, isolating the dose-response component from city and year fixed effects. Panels A and C use predictions based on actual observed hourly data for 7-day and 28-day windows, respectively. Panels B and D use predictions based on simulated data.



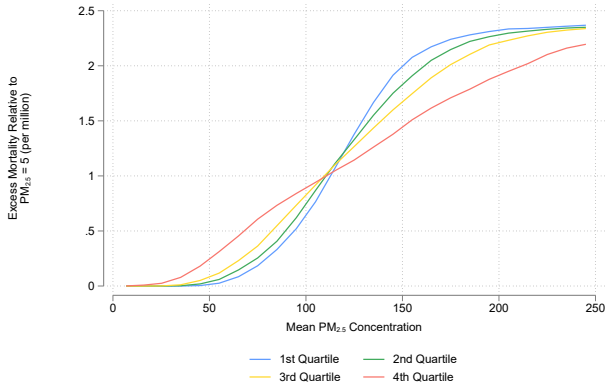
(A) 7-Day Window, Actual Data



(B) 7-Day Window, Simulated Data



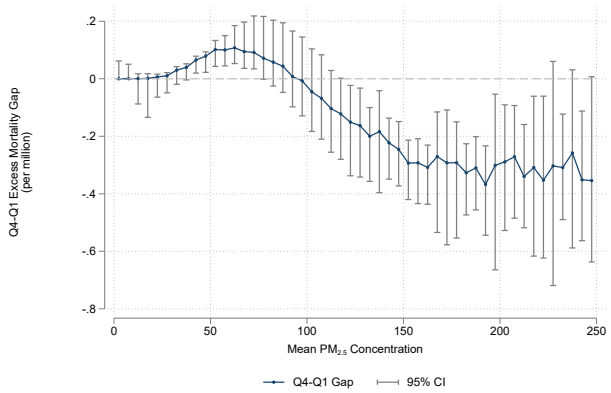
(C) 28-Day Window, Actual Data



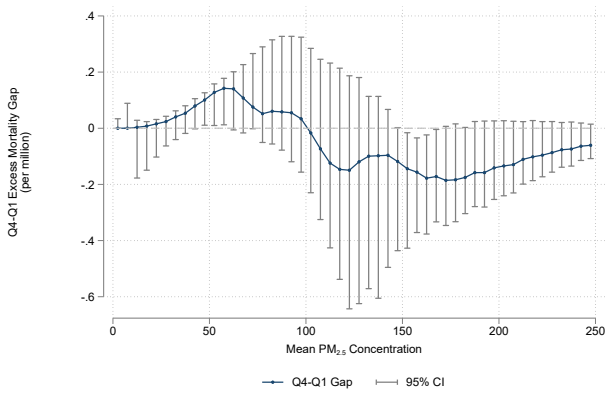
(D) 28-Day Window, Simulated Data

Figure 4: Dose-Response Curves by Quartile of Exposure Variance

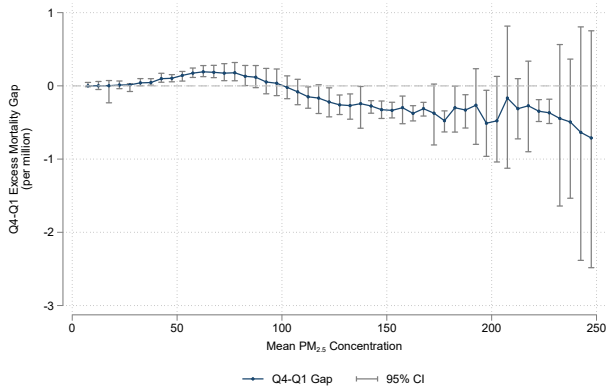
Notes: This figure shows how the dose-response relationship between mean $PM_{2.5}$ exposure and predicted mortality differs across levels of temporal exposure variance. Each colored line represents the average predicted dose-response curve for city-day observations falling into a specific quartile of exposure variance. Exposure variance is calculated for each observation as the variance of its daily average $PM_{2.5}$ concentrations over the preceding 7-day (Panels A, B) or 28-day (Panels C, D) window. The dataset is divided into four quartiles based on the overall distribution of this variance metric. Each line then plots the mean predicted mortality against the mean $PM_{2.5}$ exposure for all observations within that variance quartile. Predictions are generated using the cross-fitted regression coefficients from Equation (A2), applied to the specific distribution of hourly pollution for each city-day observation. Weather variables are held constant at their sample means. Predictions are expressed as excess mortality relative to $PM_{2.5} = 5$ $\mu g/m^3$ with constant exposure, isolating the dose-response component from city and year fixed effects. Panels A and C use actual observed data; Panels B and D use simulated data.



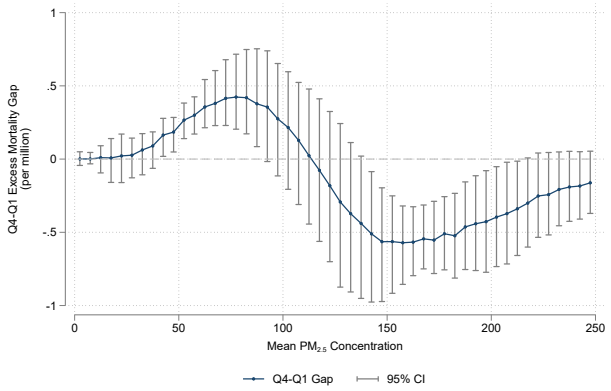
(A) 7-Day Window, Actual Data



(B) 7-Day Window, Simulated Data



(C) 28-Day Window, Actual Data



(D) 28-Day Window, Simulated Data

Figure 5: Q4–Q1 Variance-Quartile Gap in Predicted Excess Mortality, with Bootstrap 95% CI

Notes: This figure shows the difference in predicted excess mortality between the highest (Q4) and lowest (Q1) exposure variance quartiles at each mean $PM_{2.5}$ level. Connected markers plot the bootstrap median of the Q4–Q1 gap; grey whiskers show pointwise 95% confidence intervals (2.5th–97.5th percentiles across 250 bootstrap iterations). The dashed horizontal reference line marks zero. Positive values indicate the high-variance quartile (Q4) has higher predicted mortality; negative values indicate the low-variance, near-constant-exposure quartile (Q1) is more harmful. Variance-quartile assignments are fixed on the original (non-resampled) data and held constant across bootstrap draws (see Appendix Section A1.2.5). The x-axis is the mean $PM_{2.5}$ concentration ($\mu g/m^3$) over the preceding 7 days (Panels A, B) or 28 days (Panels C, D). The y-axis is the Q4–Q1 gap in excess mortality per 1 million population (relative to $PM_{2.5}$ = 5 $\mu g/m^3$ with constant exposure). Panels A and C use actual observed data; Panels B and D use simulated data. Plots are truncated at 250 $\mu g/m^3$.

References

[1] J. S. Apte, M. Brauer, A. J. Cohen, M. Ezzati, C. A. Pope, Ambient PM2.5 Reduces Global and Regional Life Expectancy. *Environmental Science & Technology Letters* **5** (9), 546–551 (2018).

[2] C. A. Pope, D. W. Dockery, Air pollution and life expectancy in China and beyond. *Proceedings of the National Academy of Sciences* **110** (32), 12861–12862 (2013).

[3] A. Ebenstein, M. Fan, M. Greenstone, G. He, M. Zhou, New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy. *Proceedings of the National Academy of Sciences* **114** (39), 10384–10389 (2017).

[4] S. S. Lim, *et al.*, A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet* **380** (9859), 2224–2260 (2012).

[5] J. Lelieveld, J. S. Evans, M. Fnais, D. Giannadaki, A. Pozzer, The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* **525** (7569), 367–371 (2015).

[6] A. J. Cohen, *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. *The Lancet* **389** (10082), 1907–1918 (2017).

[7] S. Heft-Neal, J. Burney, E. Bendavid, M. Burke, Robust relationship between air quality and infant mortality in Africa. *Nature* **559** (7713), 254–258 (2018).

[8] M. J. Daniels, F. Dominici, J. M. Samet, S. L. Zeger, Estimating particulate matter-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest US cities. *American journal of epidemiology* **152** (5), 397–406 (2000).

[9] G. Papadogeorgou, M.-A. Kioumourtzoglou, D. Braun, A. Zanobetti, Low levels of air pollution and health: effect estimates, methodological challenges, and future directions. *Current environmental health reports* **6** (3), 105–115 (2019).

[10] C. Ren, S. Tong, Health effects of ambient air pollution–recent research development and contemporary methodological challenges. *Environmental Health* **7** (1), 1–10 (2008).

[11] Q. Di, *et al.*, Air pollution and mortality in the Medicare population. *New England Journal of Medicine* **376** (26), 2513–2522 (2017).

[12] R. Burnett, *et al.*, Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proceedings of the National Academy of Sciences* **115** (38), 9592–9597 (2018).

[13] F. Liang, *et al.*, Long-term exposure to fine particulate matter and cardiovascular disease in China. *Journal of the American College of Cardiology* **75** (7), 707–717 (2020).

[14] Y. Wei, *et al.*, Emulating causal dose-response relations between air pollutants and mortality in the Medicare population. *Environmental Health* **20** (1), 1–10 (2021).

[15] Y. Gong, S. Li, N. J. Sanders, G. Shi, The mortality impact of fine particulate matter in China: Evidence from trade shocks. *Journal of Environmental Economics and Management* **117**, 102759 (2023), doi:https://doi.org/10.1016/j.jeem.2022.102759, https://www.sciencedirect.com/science/article/pii/S0095069622001127.

[16] T. Deryugina, G. Heutel, N. H. Miller, D. Molitor, J. Reif, The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review* **109** (12), 4178–4219 (2019).

[17] N. H. Miller, D. Molitor, E. Zou, *The Nonlinear Effects of Air Pollution on Health: Evidence from Wildfire Smoke*, Tech. Rep. 32924, National Bureau of Economic Research (2024), https://doi.org/10.3386/w32924.

[18] Q. Zeng, G. Li, L. Zhao, G. Jiang, X. Pan, Characteristics of the exposure-response relationship of particulate matter and mortality. *Journal of Occupational and Environmental Medicine* **57** (10), e93–e100 (2015).

[19] J.-Y. Son, M. L. Bell, The relationships between short-term exposure to particulate matter and mortality in Korea: Impact of particulate matter exposure metrics for sub-daily exposures. *Environmental Research Letters* **8** (1), 014015 (2013).

[20] H. Lin, *et al.*, Is standard deviation of daily PM2.5 concentration associated with respiratory mortality? *Environmental Pollution* **216**, 208–214 (2016), doi:https://doi.org/10.1016/j.envpol.2016.05.068, https://www.sciencedirect.com/science/article/pii/S0269749116304560.

[21] S. Heft-Neal, *et al.*, Emergency department visits respond nonlinearly to wildfire smoke. *Proceedings of the National Academy of Sciences of the United States of America* **120** (39), e2302409120 (2023), doi:10.1073/pnas.2302409120, https://doi.org/10.1073/pnas.2302409120.

[22] K. Ito, S. Zhang, Willingness to Pay for Clean Air: Evidence from Air Purifier Markets in China. *Journal of Political Economy* **128** (5), 1627–1672 (2020), doi:10.1086/705554, https://doi.org/10.1086/705554.

[23] R. T. Burnett, *et al.*, An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environmental health perspectives* **122** (4), 397–403 (2014).

[24] M. M. Nasari, *et al.*, 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 & Health* **9** (8), 961–972 (2016).

[25] L. D. Hill, *et al.*, Health assessment of future PM2. 5 exposures from indoor, outdoor, and

secondhand tobacco smoke concentrations under alternative policy pathways in Ulaanbaatar, Mongolia. *PLoS One* **12** (10), e0186834 (2017).

[26] L. Forastiere, M. Carugno, M. Baccini, Assessing short-term impact of PM 10 on mortality using a semiparametric generalized propensity score approach. *Environmental Health* **19** (1), 1–13 (2020).

[27] P. J. Landrigan, *et al.*, The Lancet Commission on Pollution and Health. *The Lancet* **391** (10119), 462–512 (2018).

[28] U.S. Environmental Protection Agency, *Supplement to the 2019 Integrated Science Assessment for Particulate Matter (Final Report, 2022)*, Tech. Rep. EPA/635/R-22/028, U.S. Environmental Protection Agency, Washington, DC (2022).

[29] M. Greenstone, G. He, R. Jia, T. Liu, Can technology solve the principal-agent problem? Evidence from China's war on air pollution. *American Economic Review: Insights* **4** (1), 54–70 (2022).

[30] M. H. Forouzanfar, *et al.*, 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* **388** (10053), 1659–1724 (2016).

[31] O. Adetona, *et al.*, Review of the health effects of wildland fire smoke on wildland firefighters and the public. *Inhalation toxicology* **28** (3), 95–139 (2016).

[32] J. A. Estévez-García, N. Y. Rojas-Roa, A. I. Rodríguez-Pulido, Occupational exposure to air pollutants: particulate matter and respiratory symptoms affecting traffic-police in Bogotá. *Revista de salud publica* **15** (6), 870–885 (2013).

[33] I. Gutiérrez-Avila, R. O. Wright, M. J. Rosa, A. C. Just, Exposure to daily mean and maximum 1-hour $PM_{2.5}$ concentrations and pediatric respiratory mortality in the Mexico

City Metropolitan Area. *Environmental Epidemiology* **9** (4), e408 (2025), doi:10.1097/EE9.0000000000000408.

[34] T. Chen, *et al.*, Xgboost: extreme gradient boosting. *R package version 0.4-2* **1** (4), 1–4 (2015).

Acknowledgments

We are grateful to David Molitor, Eran Bendavid, Carlos Gould, Steve Luby, and seminar participants at Stanford University for helpful comments and discussions. We also thank Bo Lin, Jack Shane, and Liyu Zheng for excellent research assistance.

Author contributions: Conceptualization: BG, KSB, HL, LM, LH, GM
Methodology: BG, KSB, HH, GM, AS, AM
Data Acquisition: HL, LM, PY, MZ
Writing: BG, KSB, HL, LM, LH, HH, GM, AS, AM
Funding Acquisition: BG, KSB, HL, LH, GM
Formal Analysis: BG, KSB, AS, MW
Supervision: GM

Competing interests: There are no competing interests to declare.

Data and materials availability: The hourly $PM_{2.5}$ concentration measures for the 251 cities, the meteorological variables, the simulation outputs, and all program files that generate the tables and figures in this paper will be deposited in a public repository upon publication and can be shared with editors and referees upon request. The daily mortality counts were obtained from the Chinese Center for Disease Control and Prevention's Disease Surveillance Points (DSP) system. These mortality microdata are confidential and cannot be redistributed; they are available upon request to and at the discretion of the Chinese Center for Disease Control and Prevention, and any analyses using these data must be carried out on the Center's secure server.

Supplementary Materials for: Beyond Mean Exposure: How Exposure Variance Shapes Air Pollution Mortality Risk

Bissan Ghaddar[1,†], Kimberly Singer Babiarz[2,†], Hongbin Li[2], Lingsheng Meng[2], Lynn Hildemann[2], Han Hong[2], Alvise Scarabosio[2], Meili Wang[3], Peng Yin[4], Maigeng Zhou[4], Grant Miller[2,5,*]

[*]Corresponding author. Email: ngmiller@stanford.edu

[†]These authors contributed equally to this work.

This PDF file includes:

A1 Data and Methods

A1.1 Data

In this paper, we use hourly air quality data linked to daily vital statistics from 251 Chinese cities over a period of four years (2013-16) to generate new estimates of the dose-response curve relating air pollution to mortality. Our data sets are novel, large-scale, and link three different data sources. First, we match this city-day mortality data with city-hour $PM_{2.5}$ air pollution concentration measurements obtained through the National Urban Air Quality Real-time Publishing Platform. This data contains hourly concentration measurements of $PM_{2.5}$ in 251 cities from 2013 through 2016 and has generally been found to be good quality (29). Second, we use city-day mortality data from the Chinese Center for Disease Control and Prevention's Disease Surveillance Points system. With this data, we measure city-level daily mortality rates per million population due to diseases sensitive to air quality (lung disease deaths (pneumonia, asthma, chronic obstructive pulmonary disease) and cardiovascular disease deaths) (30). Third, we match this linked data with city-day information from China's Meteorological Data Sharing Service System on meteorological conditions (mean temperature, mean relative humidity, and mean air pressure [4]).

To our knowledge, no previous studies have used hourly ambient $PM_{2.5}$ air pollution data to study mortality in large, geographically-dispersed populations. Those studying mortality in larger populations have focused on subgroups (e.g., Medicare enrollees, firefighters, traffic officers) (16; 31; 32), focused on single metropolitan areas (33), or have used PM_{10} rather than $PM_{2.5}$ (19).

This linked dataset has at least two advantages. The first is that its high-frequency air pollution measurements (hourly) enable us to study how distributions of pollution exposure (rather than annual, or even daily, means) are related to mortality risk. The second is that our data contains a substantial number of very high pollution concentration measurements from geographically diverse

[4]Missing values and outliers are not observed in the mortality or meteorological data sets, and no preprocessing measures were applied. Hourly pollution concentration data is processed to eliminate a small number of outlier readings. Extremely high pollution values inconsistent with adjacent values, or values equal to exactly zero suggest monitor malfunction and are eliminated from the dataset. A total of 898 hourly readings are dropped from 6.08 million records (0.015%).

areas of China – with 1.2 million hourly $PM_{2.5}$ measurements above 75 $\mu g/m^3$ (or an AQI 161) up to a maximum observed concentration of 861.7 $\mu g/m^3$. Appendix Figures A2 and A3 show the location and distribution of pollution measurements in our sample across space and over time. Appendix Figure A1 also shows the distribution of share of hours spent across $PM_{2.5}$ concentration bins (0-10, 10-20, ..., 440-450 $\mu g/m^3$) for the actual and simulated samples of city-days over the preceding 7 and 28 days.

A1.2 Methods

With our linked data, we use a five-step approach to estimate the pollution-mortality dose-response curve. (i) We split the 251 cities in our sample into two folds and use a cross-fitting procedure to keep threshold identification and dose-response estimation on non-overlapping subsets of cities. (ii) On each training fold, we use XGBoost to identify concentration thresholds and define discrete $PM_{2.5}$ bins. (iii) On the held-out fold, we estimate the dose-response relationship by OLS using the share of hours each city-day spends in those bins. (iv) We use the resulting OLS estimates to predict daily mortality on both observed and simulated $PM_{2.5}$ concentration values, tracing out dose-response curves across mean and variance combinations. (v) We bootstrap-resample cities within folds to obtain pointwise confidence intervals for the resulting dose-response curves.

A1.2.1 Cross-Fitting and Prediction Stacking

Because the same city-day observations are used both to identify concentration thresholds (step ii) and to estimate the dose-response coefficients (step iii), using the full sample for both could introduce overfitting bias. To avoid this, we wrap steps (ii) and (iii) in the following cross-fitting procedure. Cities are assigned to Fold A (125 cities) or Fold B (126 cities) by a single random draw; the assignment is preserved across all subsequent steps (including bootstrap iterations). For each direction, XGBoost is applied to its training fold to identify concentration thresholds as described in Section A1.2.2, and OLS (Equation (A2)) is then estimated on the held-out fold using the thresholds derived from the training fold. Predicted excess mortality for a given city is always computed using

the coefficients and thresholds from the direction that did not include that city in training. This implies that predictions for Fold A cities use Direction B, and predictions for Fold B cities use Direction A. We concatenate these non-overlapping predictions to obtain full-sample coverage while eliminating the overfitting bias at the threshold-selection step that would otherwise arise from predicting on training data. Hence, each city's regressors are constructed using thresholds chosen from a non-overlapping set of cities. Moreover, bootstrap resampling (Appendix Section A1.2.5) quantifies the resulting OLS sampling uncertainty. For simulated data (which is not attached to any fold), we average predictions across the two directions.

A1.2.2 Identification of Thresholds in the Dose-Response Curve

To measure cumulative share of hours spent in different $PM_{2.5}$ bins over 7- and 28-day periods (i.e., periods of one week and approximately one month), we use a machine learning model to help us identify the thresholds, or changes in shape, in the pollution-mortality relationship that are most predictive of mortality across the full distribution of air pollution measurements. We use a machine learning model for this task to be flexible and capture the non-linear relationship between pollution and mortality. Our machine learning approach automates the discovery of these complex patterns, handling the scale and complexity of the data. We use these thresholds to define discrete concentration intervals over which we then compute the cumulative share of hours spent during the preceding 7 and 28 days. The procedure described below is applied separately on each training fold (Fold A and Fold B) under the cross-fitting wrapper of Section A1.2.1.

In total, our model incorporates 271 features, including the total share of hours within the preceding week (7 days) or month (28 days) during which $PM_{2.5}$ concentrations were measured above each observed incremental 10 $\mu g/m^3$ value (i.e., the share of time spent above 10 $\mu g/m^3$, above 20 $\mu g/m^3$, $\cdots$), continuous measure of daily average temperature, daily average humidity, city indicators, and year indicators. From these, we identify a set of $G-1$ thresholds $\{\tau_1 < \tau_2 < \cdots < \tau_{G-1}\}$, below 150 $\mu g/m^3$, from a candidate grid of shares of time above $10, 20, \ldots, 150$ $\mu g/m^3$ using XGBoost. These thresholds define G non-overlapping concentration intervals or

bins with lower and upper bounds $(c^g_{\min}, c^g_{\max}]$ for $g \in \{1, \ldots, G\}$ (with $c^1_{\min} = 0$ and $c^G_{\max} = \infty$ by convention). XGBoost, a decision-tree-based ensemble machine learning algorithm that uses a gradient boosting framework, ranks candidate threshold features by their contribution to mortality prediction, accounting for weather conditions, city effects, and year fixed effects. XGBoost hyperparameters are tuned once per fold via 5-fold cross-validation within that fold, and held fixed throughout the analysis, including across bootstrap iterations.

More precisely, we first partition our data into low pollution ($PM_{2.5} \leq 35.5\,\mu g/m^3$) and high pollution ($PM_{2.5} > 35.5\,\mu g/m^3$) groups based on the mean exposure during the observation window[5]. We then apply XGBoost separately to each pollution group to identify the most important concentration thresholds within each regime, recognizing that optimal partitions may differ across pollution levels. From each group, we extract the top 3 most important threshold features and combine them to form our final set of concentration bins, applying a minimum gap constraint of 20 $\mu g/m^3$ between consecutive thresholds to ensure meaningful bin widths. The resulting set of candidate thresholds is then duplicated across the two pollution groups and passed through the 20 $\mu g/m^3$ minimum-gap filter, so the total number of surviving thresholds is typically below six. In practice, the spectrum of $PM_{2.5}$ concentration is partitioned such that the within-partition variation in mortality is minimized and cross-partition differences are maximized using the XGBoost algorithm (34).

In our main model, three to four thresholds were chosen per fold-window, all at or above 70 $\mu g/m^3$. In fact, even within the low-pollution partition, the most predictive threshold features for mortality are themselves elevated, because within-window variation captured by XGBoost appears to be driven by sporadic hours of high exposure rather than by modest fluctuations around lower means. Because XGBoost is applied separately on each fold through our cross-fitting procedure, the identified thresholds differ between folds. For the 7-day window, Fold A yields thresholds of 110, 130, and 150 $\mu g/m^3$ (and thus the four bins $(0, 110]$, $(110, 130]$, $(130, 150]$, and >150) while Fold B yields 70, 100, 130, and 150 $\mu g/m^3$ (bins $(0, 70]$, $(70, 100]$, $(100, 130]$, $(130, 150]$, >150). For

[5] 35.5 $\mu g/m^3$ is the threshold where AQI is 100.

the 28-day window, Fold A yields 80, 100, 130, and 150 $\mu g/m^3$ (bins $(0, 80]$, $(80, 100]$, $(100, 130]$, $(130, 150]$, >150) and Fold B yields 80, 130, and 150 $\mu g/m^3$ (bins $(0, 80]$, $(80, 130]$, $(130, 150]$, >150).

For each interval or bin, we then compute the share of hours measured in that bin over the preceding 7 or 28 days for each city-day observation:

$$D_{jt}^{g} = \frac{1}{H_{jt}} \sum_{h=1}^{H_{jt}} \mathbf{1}(c_{\min}^{g} < c_{jt,h}^{pm2.5} \leq c_{\max}^{g}), \tag{A1}$$

where H_{jt} is the total number of hours in the reference period preceding day t (168 for the 7-day window, 672 for the 28-day window), $c_{\min}^{g}$ and $c_{\max}^{g}$ are the lower and upper bounds of bin g, and $c_{jt,h}^{pm2.5}$ is the $PM_{2.5}$ concentration in city j at the h-th hour of that reference period.

A1.2.3 Estimation of the Dose-Response Relationship

We next use these hourly concentration duration intervals for each city-day observation to parameterize a more parsimonious Ordinary Least-Squares (OLS) model to estimate the dose-response relationship between $PM_{2.5}$ exposure and mortality:

$$M_{jt} = \sum_{g=1}^{G} \beta_g D_{jt}^{g} + \gamma W_{jt} + \delta Y_t + \zeta C_j + \epsilon_{jt} \tag{A2}$$

where M_{jt} is the number of deaths per 1,000,000 persons in city j on day t, D_{jt} is the share of hours when exposures to $PM_{2.5}$ are in a given range g during the 7 or 28 day period preceding day t in city j, Y_t denotes the fixed effects for year, C_j is the fixed effects for city j, W_{jt} is daily mean temperature, humidity, and air pressure in city j on day t. Because the interval or bin definitions D_{jt}^{g} depend on thresholds identified by XGBoost, the cross-fitting procedure ensures that the thresholds used to construct the regressors for a given city were not estimated using that city's data. Standard errors are clustered at the city level.

A1.2.4 Simulation and Visualization of How the Dose-Response Curve Depends on Exposure Variance

Finally, we use these OLS model estimates in two simulation exercises to trace out the full dose-response curves for the relationship between $PM_{2.5}$ exposure and mortality across a wide range of mean/variance exposure combinations. Separately for 7- and 28-day periods, we predict daily mortality using both (1) values of $PM_{2.5}$ concentrations as observed in our data, and (2) simulated values of $PM_{2.5}$ concentrations.

For the simulated $PM_{2.5}$ concentration values, we generate a synthetic dataset allowing us to explore combinations of hourly exposure, mean, and variance not present in the observed data. For each integer target concentration from 1 to 499 $\mu g/m^3$, we draw 100 daily means from a Beta distribution centered at the target, with a day-to-day variance level (low, medium, or high) assigned once per target. For each daily-mean draw we generate 10 replicate synthetic city-days, all sharing that daily mean. Each replicate is summarized by 5 values drawn from a normal distribution centered at the daily mean with standard deviation 10%, 20%, or 30% of the mean (with probabilities 0.3, 0.4, 0.3, respectively). These 5 values are a discretization used to compute the share-of-time-above-threshold features at each 10 $\mu g/m^3$ increment. The resulting observations are concatenated in order of increasing target concentration, and 7- and 28-day trailing moving averages (and corresponding day-to-day variance measures) are computed over the ordered list of simulated days (treating the concatenated observations as a synthetic time series).[6]

A1.2.5 Bootstrap Confidence Intervals

We construct pointwise bootstrap confidence intervals for the dose-response curve (Figure 1) and for the Q4–Q1 variance-quartile gap (Figure 5). For each of 250 bootstrap iterations, we resample cities with replacement within each fold (preserving the Fold A / Fold B assignment from the point-estimate pipeline) and re-run the full cross-fitting procedure: XGBoost is trained on the resampled

[6]The simulation generates approximately 499,000 observations (499 target concentrations × 100 daily means × 10 replicates).

fold to identify thresholds, and OLS is estimated on the opposite fold using those thresholds. This procedure is performed in both directions, yielding two coefficient vectors per iteration. XGBoost hyperparameters are held fixed at their Section A1.2.1 values across all iterations.

For each iteration b, we predict excess mortality at each $PM_{2.5}$ level x relative to a reference of $PM_{2.5} = 5\,\mu g/m^3$ with constant exposure:

$$\Delta^{(b)}(x) = \sum_{g=1}^{G} \beta_g^{(b)} D_g^{(b)}(x) \; - \; \beta_{\text{first}}^{(b)} \times 100, \tag{A3}$$

where $\beta_{\text{first}}^{(b)}$ is the coefficient on the lowest concentration bin in iteration b. At the reference level, 100% of hours fall in the lowest bin, so $\Delta^{(b)}(5) = 0$ by construction. Because the reference prediction depends only on $\beta_{\text{first}}^{(b)}$ and the (fixed) sample-mean weather constant, this differencing cancels the intercept, the weather term, and the city and year fixed-effect contributions within each bootstrap draw. The reported point estimate at each $PM_{2.5}$ level is $\Delta(x)$ from the original (non-bootstrap) cross-fitted estimates. The pointwise 95% confidence interval is the 2.5th–97.5th percentile of $\{\Delta^{(1)}(x), \ldots, \Delta^{(250)}(x)\}$.

For the variance-quartile gap, variance-quartile assignments are made once on the original (non-resampled) data and held fixed across all 250 bootstrap draws. For each draw b and each quartile $q \in \{1, 4\}$, the quartile curve is the mean of $m_{\text{pred}}^{(b)}(x)$ over observations assigned to Q_q; the gap at each $PM_{2.5}$ level is $\text{gap}^{(b)}(x) = \Delta_{Q_4}^{(b)}(x) - \Delta_{Q_1}^{(b)}(x)$, and its pointwise 95% confidence interval is the 2.5th–97.5th percentile across draws.

A2 Robustness

We investigate the robustness of our results in three ways. First, we consider the extent to which our approach to partitioning the $PM_{2.5}$ distribution in our gradient boost procedure influences the shape of the dose-response curve. Our approach creates categorical variables measuring the share of hours with $PM_{2.5}$ exposure up to $150\,\mu g/m^3$ as separate features, with all exposure above $150\,\mu g/m^3$ combined into a single exposure category. This choice is motivated by the relative

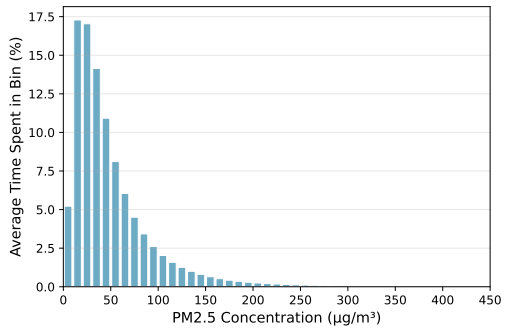
sparsity of observations at very high pollution levels. However, this approach also potentially creates a 'ceiling effect' in the estimated relationship between $PM_{2.5}$ exposure and mortality at high $PM_{2.5}$ concentrations. We highlight that even the presence of such a ceiling effect would not change the substantial variability in mortality that we find at each mean $PM_{2.5}$ value below 150 $\mu g/m^3$. Nonetheless, to investigate the possibility that a ceiling effect is an artifact of interval coding, we repeat our exercise using higher truncation points of 200 $\mu g/m^3$ and 250 $\mu g/m^3$ in the $PM_{2.5}$ distribution and re-estimate dose-response curves using simulated city-hour observations. Appendix Figure A4 shows the resulting curves. In general, they share similar properties to those in Figure 2: the upper bounds of all curves are generally similar to each other, and all curves are convex at lower average $PM_{2.5}$ exposures and mostly concave at higher levels of exposure (above about 145 $\mu g/m^3$). The estimates do become noisier at the highest concentrations. Overall, however, the qualitative transition from convex at low concentrations to concave at higher concentrations is preserved across specifications, suggesting that our results at very high levels of exposure are not simply an artifact of our approach.

Second, given the substantial variability in mortality that we find at each level of $PM_{2.5}$ exposure, we assess the sensitivity of our results to directly conditioning on mean 7- and 28-day exposure in our analyses. Appendix Figure A5 shows these results, generally demonstrating that our results are not highly sensitive to conditioning directly on mean exposure.

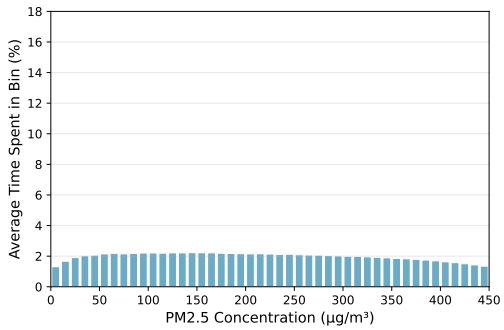
Third, we examine the extent to which seasonal patterns in pollution and mortality could drive our results by re-estimating Equation (A2) with month fixed effects added to the year and city fixed effects. Appendix Figure A6 shows the resulting dose-response curves. Their shape and magnitude closely track the baseline specification, indicating that our findings are not driven by within-year

seasonal confounding.[7]

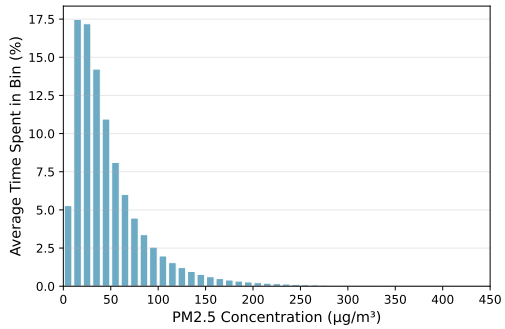
[7]Note that the 28-day curves dip slightly below the low-pollution reference at moderate concentrations (roughly 40 µg to 90 µg/m^3 in the actual data) when we include month fixed effects. The central (mean) dose-response line dips to about −0.15 (actual data) and −0.22 (simulated data) deaths per million per day relative to a peak of about 1.0 (actual data; 1.35 simulated) per million per day, with shaded prediction regions dipping further. This reflects a binning artifact rather than a true decline in risk because one cross-fitted fold isolates a narrow [80, 100) PM2.5 exposure bin for which the corresponding regression coefficient falls below the [0, 80) reference (cluster-robust Wald test of equality, $t = 3.71$, $p \approx 3 \times 10^{-4}$). As a result, some individual city-days weighted heavily in that bin have negative predicted values. The other cross-fitted fold splits the same range at PM2.5 80 and 130 and shows no such significant decline ($t = 1.40$, $p = 0.16$), demonstrating that this feature is not stable across folds. The convex-then-concave shape of the dose-response curve remains unaffected.



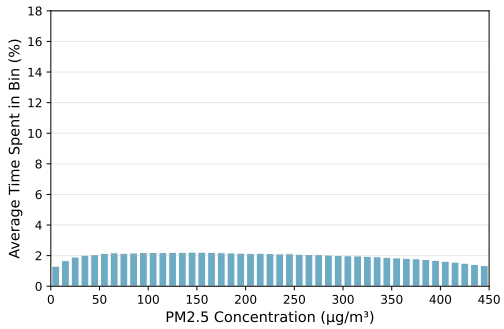
(A) 7-Day Window, Actual Data



(B) 7-Day Window, Simulated Data



(C) 28-Day Window, Actual Data



(D) 28-Day Window, Simulated Data

Figure A1: Distribution of Time Share in $PM_{2.5}$ Concentration Bins

Notes: This figure shows the average distribution of time spent across $PM_{2.5}$ concentration bins. The x-axis shows 10-$\mu g/m^3$-wide $PM_{2.5}$ concentration bins. The y-axis shows the average share of hours that city-day observations spent in each bin over the preceding 7-day (Panels A, B) or 28-day (Panels C, D) window. Panels A and C show this distribution for the actual observed data. Panels B and D show the equivalent distribution for the simulated data. Data source fo the actual data: China National Urban Air Quality Real-time Publishing Platform, Beijing.

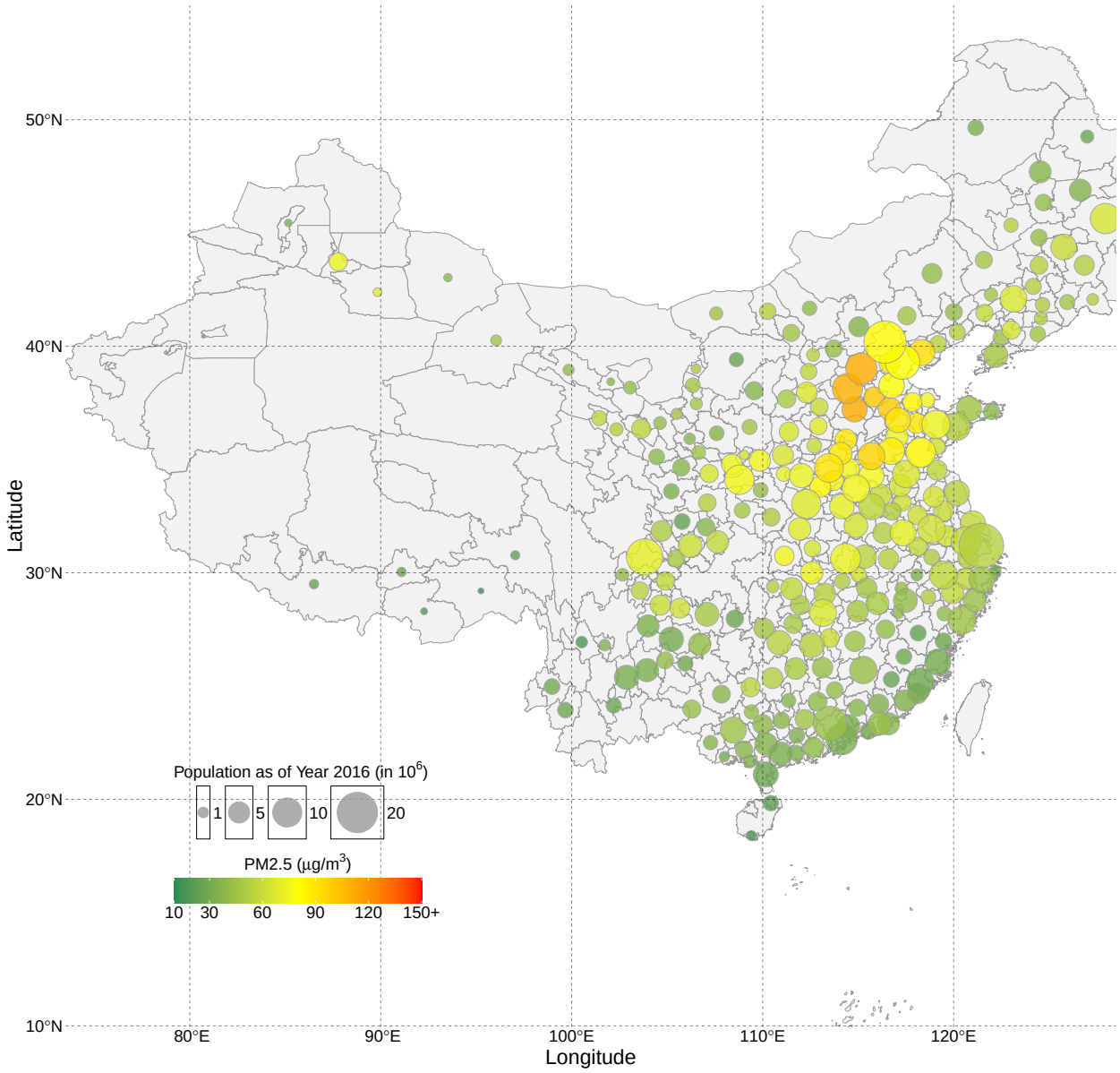


Figure A2: Geographic distribution of $PM_{2.5}$ pollution concentration.

Notes: This figure shows the geographic distribution of mean $PM_{2.5}$ concentration across the 251 Chinese cities in our sample. The value for each city is the average of all its hourly $PM_{2.5}$ measurements ($\mu g/m^3$) over the entire study period (2013-2016). Darker shades of red indicate higher average pollution levels. The map illustrates the substantial geographic variation in long-term pollution exposure, with the highest concentrations centered in the North China Plain. Data source: China National Urban Air Quality Real-time Publishing Platform.

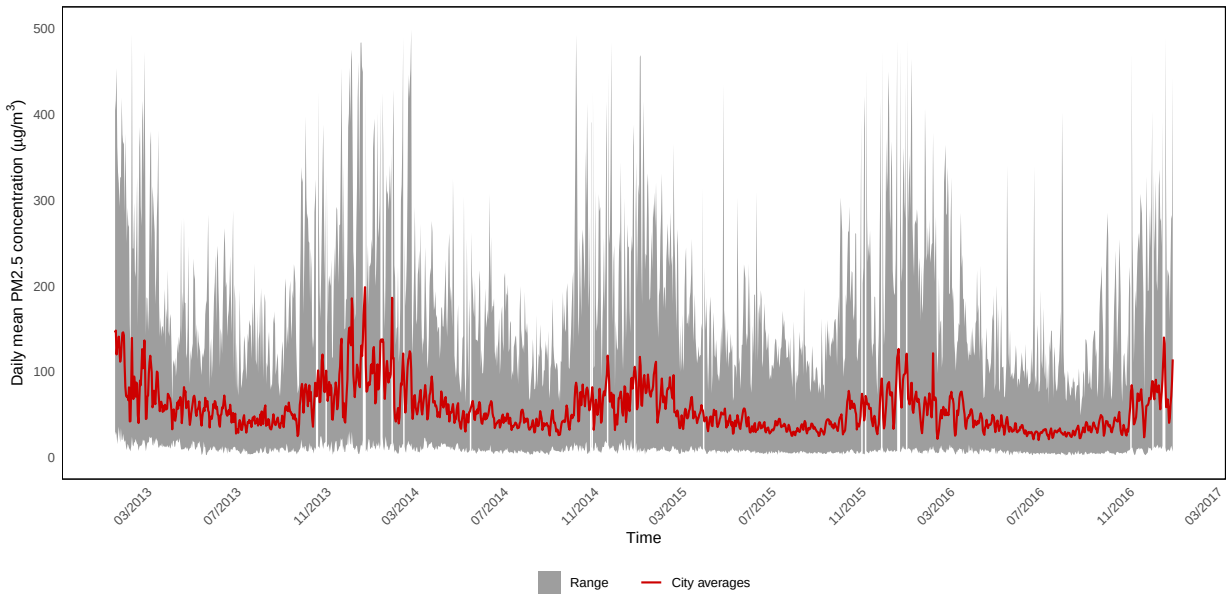
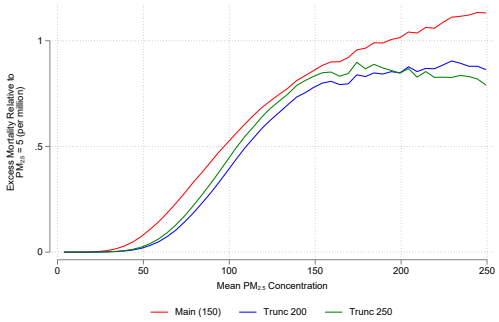
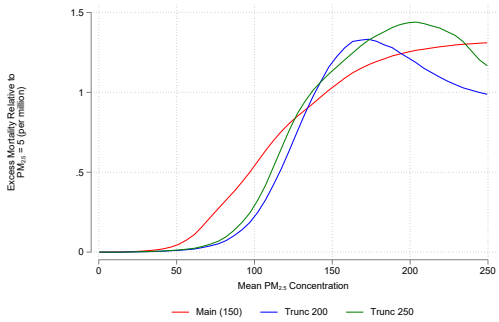


Figure A3: Temporal distribution of $PM_{2.5}$ pollution concentration.

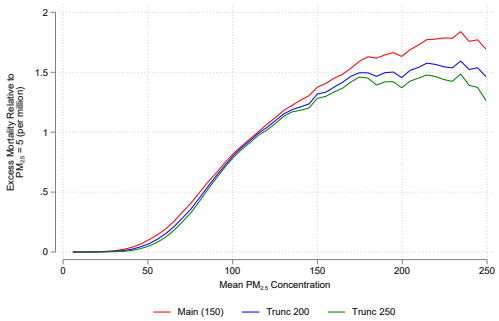
Notes: This figure plots the temporal distribution of daily mean $PM_{2.5}$ concentrations across all 251 cities from 2013 to 2016. For each day in the sample, the solid red line represents the cross-city average of daily mean $PM_{2.5}$ concentrations. The gray shaded area shows the range (minimum to maximum) of daily mean $PM_{2.5}$ concentrations observed across all cities for that day. Daily means are calculated by averaging the 24 hourly measurements for each city-day observation. The figure highlights the strong seasonality in pollution levels, with peaks during winter months, as well as the wide daily variation in pollution exposure across different locations. Data source: China National Urban Air Quality Real-time Publishing Platform.



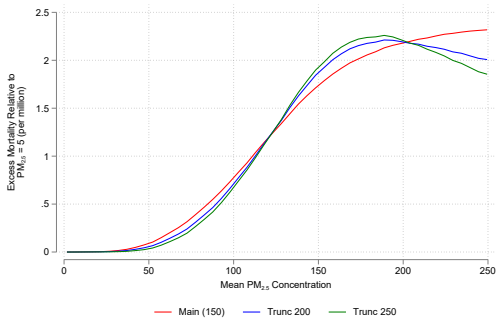
(A) 7-Day Window, Actual Data



(B) 7-Day Window, Simulated Data



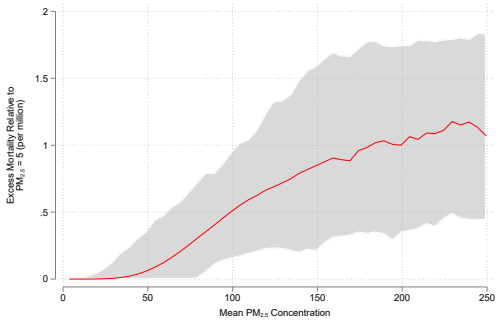
(C) 28-Day Window, Actual Data



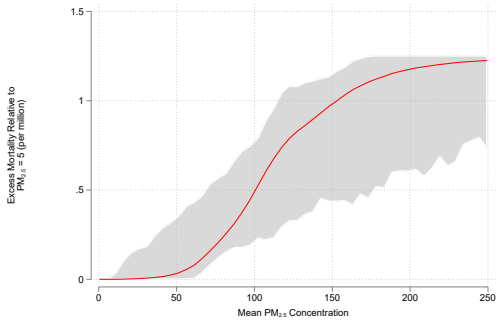
(D) 28-Day Window, Simulated Data

Figure A4: Predicted Mortality from $PM_{2.5}$ Exposure: Robustness to Different Truncation Levels

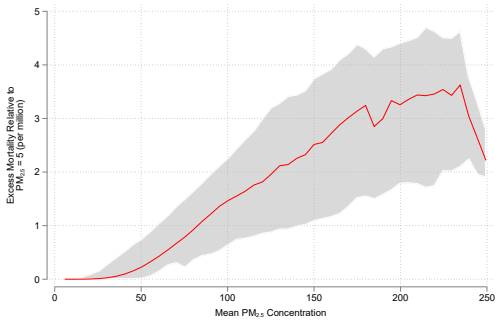
Notes: This figure presents a robustness check of the dose-response curves shown in Figure 2, testing their sensitivity to the top-coding of pollution bins in our feature selection procedure. The figure overlays the predicted dose-response curves from multiple models: the main specification (which uses bins top-coded at 150 $\mu g/m^3$) and two alternative specifications with higher top-codes (200 and 250 $\mu g/m^3$). For each model, the solid line represents the mean predicted mortality and the corresponding shaded region shows the full range of predictions for a given mean exposure. The x-axis represents the mean $PM_{2.5}$ concentration ($\mu g/m^3$) over the preceding 7 days (Panels A, B) or 28 days (Panels C, D). The y-axis is excess predicted daily mortality per 1 million population relative to a reference exposure of $PM_{2.5}$ = 5 $\mu g/m^3$ with constant exposure. Predictions for each model are generated using its respective cross-fitted regression coefficients from Equation (A2). Weather variables are held constant at their sample means. Expressing predictions as excess mortality relative to the low-pollution reference isolates the dose-response component from city and year fixed effects. Panels A and C use the actual distribution of hourly $PM_{2.5}$ concentrations observed in our sample; Panels B and D use simulated concentrations. Plots are truncated at 250 $\mu g/m^3$.



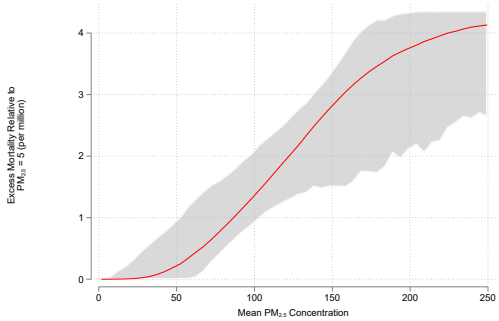
(A) 7-Day Window, Actual Data



(B) 7-Day Window, Simulated Data



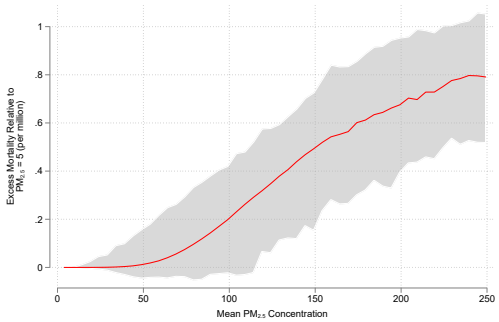
(C) 28-Day Window, Actual Data



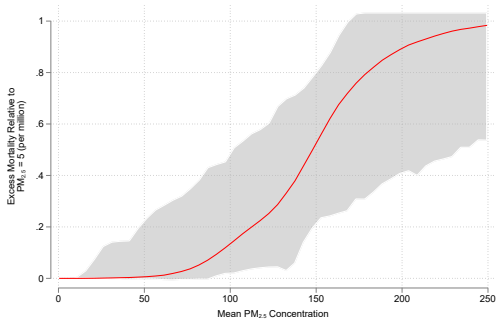
(D) 28-Day Window, Simulated Data

Figure A5: Predicted Mortality from $PM_{2.5}$ Exposure Controlling for Mean Daily Pollution

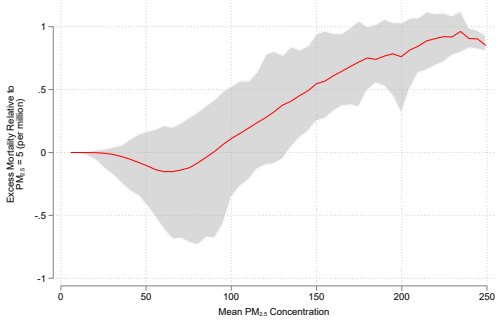
Notes: This figure presents a robustness check of the dose-response curves shown in Figure 2. The solid red line shows the mean predicted excess mortality for each level of average $PM_{2.5}$ exposure. The shaded gray region represents the full range (minimum to maximum) of predicted excess mortality values arising from all underlying hourly exposure distributions that result in the same mean $PM_{2.5}$ exposure. The x-axis represents the mean $PM_{2.5}$ concentration ($\mu g/m^3$) over the preceding 7 days (Panels A, B) or 28 days (Panels C, D). The y-axis is excess predicted daily mortality per 1 million population relative to a reference exposure of $PM_{2.5}$ = 5 $\mu g/m^3$ with constant exposure. Predictions are generated using the cross-fitted regression coefficients from Equation (A2), augmented by including a direct control for the mean $PM_{2.5}$ exposure over the relevant window. Weather variables are held constant at their sample means. Expressing predictions as excess mortality relative to the low-pollution reference isolates the dose-response component from city and year fixed effects. Panels A and C use the actual distribution of hourly $PM_{2.5}$ concentrations observed in our sample. Panels B and D use simulated concentrations. Plots are truncated at 250 $\mu g/m^3$ due to data sparsity.



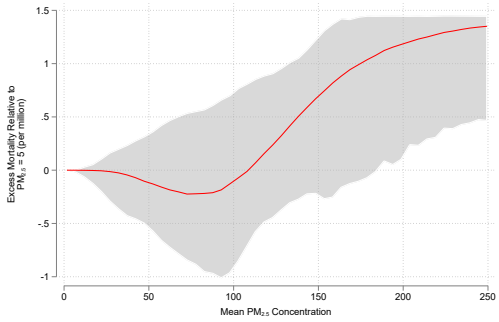
(A) 7-Day Window, Actual Data



(B) 7-Day Window, Simulated Data



(C) 28-Day Window, Actual Data



(D) 28-Day Window, Simulated Data

Figure A6: Predicted Mortality from $PM_{2.5}$ Exposure Controlling for Month Fixed Effects

Notes: This figure presents a robustness check of the dose-response curves shown in Figure 2. The solid red line shows the mean predicted excess mortality for each level of average $PM_{2.5}$ exposure. The shaded gray region represents the full range (minimum to maximum) of predicted excess mortality values arising from all underlying hourly exposure distributions that result in the same mean $PM_{2.5}$ exposure. The x-axis represents the mean $PM_{2.5}$ concentration ($\mu g/m^3$) over the preceding 7 days (Panels A, B) or 28 days (Panels C, D). The y-axis is excess predicted daily mortality per 1 million population relative to a reference exposure of $PM_{2.5}$ = 5 $\mu g/m^3$ with constant exposure. Predictions are generated using the cross-fitted regression coefficients from Equation (A2), augmented by including month fixed effects in addition to the city and year fixed effects of the baseline specification. Weather variables are held constant at their sample means. Expressing predictions as excess mortality relative to the low-pollution reference isolates the dose-response component from city, year, and month fixed effects. Panels A and C use the actual distribution of hourly $PM_{2.5}$ concentrations observed in our sample. Panels B and D use simulated concentrations. Plots are truncated at 250 $\mu g/m^3$ due to data sparsity.