<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessed</td>
<td>November 6, 2017 3:38:59 PM EST</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:32644534">http://nrs.harvard.edu/urn-3:HUL.InstRepos:32644534</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>

(Article begins on next page)
ESTIMATING HEALTH EFFECTS OF TEMPERATURE AND PM2.5 USING SATELLITE-RETRIEVED HIGH-RESOLUTION EXPOSURES

LIUHUA SHI

A Dissertation Submitted to the Faculty of the Harvard T.H. Chan School of Public Health in Partial Fulfillment of the Requirements for the Degree of Doctor of Science in the Department of Environmental Health Harvard University

Boston, Massachusetts.

November, 2016
Dissertation Advisor: Dr. Joel Schwartz

Liuhua Shi

Estimating Health Effects of Temperature and PM$_{2.5}$ Using Satellite-Retrieved High-Resolution Exposures

Abstract

Human activities emit greenhouse gases (GHGs) and air pollutants, which would affect the environment and in turn affect human health. Accurate estimate of the health effects requires high resolution exposure data of environmental stressors, such as air temperature ($T_a$) and fine particulate matter (PM$_{2.5}$). The availability of those exposure data, however, is usually limited by sparsely distributed ground monitoring network.

Therefore, the first chapter estimates $T_a$ at a fine scale on a daily basis by incorporating satellite-based remote sensing data. Satellite can provide a global daily estimate of 1 km × 1 km surface temperature ($T_s$), which is correlated with $T_a$. Hence, a statistical calibration approach between $T_a$ and $T_s$ was used to retrieve daily mean $T_a$ at 1 km resolution for the Southeastern United States for the years 2000 to 2014.

The second chapter investigates the chronic effects of temperature and its variability on mortality in New England, by using the satellite-retrieved daily mean $T_a$ estimated from previous studies. Our findings indicate that the variability of atmospheric temperature emerges as a key factor of the potential health impacts of climate change.

The last chapter examines the association between low-concentration PM$_{2.5}$ and mortality in New England, by using the satellite-retrieved PM$_{2.5}$ estimates. Our findings suggest that adverse health effects occur at low levels of fine particles, even for levels not exceeding the newly revised EPA standards.
# Table of Contents

Abstract ........................................................................................................................................... ii
Table of Contents .......................................................................................................................... iii
List of Figures with Captions .......................................................................................................... iv
List of Tables with Captions .......................................................................................................... vi
Acknowledgements ...................................................................................................................... viii
Previously Published Work ........................................................................................................ iii
Chapter I: Estimating daily air temperature across the Southeastern United States using high-resolution satellite data: a statistical modeling study ........................................ 1
Chapter II: Impacts of temperature and its variability on mortality in New England .............. 29
Chapter III: Low-concentration PM2.5 and mortality: Estimating acute and chronic effects in a population based study ........................................................................... 53
List of Figures with Captions

Chapter I: Estimating daily air temperature across the Southeastern United States using high-resolution satellite data: a statistical modeling study

Figure 1  Map of the study area showing all available NCDC air temperature monitor stations across southeastern USA for 2000-2014........................................6

Figure 2  Scatter plots of the monitored air temperature versus satellite surface temperature (left) and the monitored air temperature versus that from the stage 1 calibration (right). Data are shown for the year 2011.........................14

Figure 3  Predicted air temperature (°C) from stage 2 (top) and both stage 2 & 3 (bottom) in each 1 km × 1 km grid on a sample day (2011.08.25) across the southeastern USA..........................................................17

Figure 4  Spatial pattern of predicted air temperature (°C), averaged over the 2011 for the southeastern USA..............................................................20

Figure 5  Predicted air temperature (°C) at a 1 km × 1 km grid for August 3, 2011 (top) and August 25, 2011 (bottom) .........................................................21

Figure 6  A scatter plot of the calculated daily average $T_a$ versus the daily $T_{2m}$ from MERRA. Data are shown for the year 2011.................................22

Chapter II: Impacts of Temperature and its Variability on Mortality in New England
Figure 1 Percent increase in mortality for per 1 °C and per Interquartile Range (IQR) increases in seasonal mean temperature and temperature variability across New England, 2003-2008

Figure 2 Modifications of the effect for annual anomalies of different temperature indices by population density, age, sex, and race

Figure 3 Modifications of the effect for spatial variability in different temperature indices by population density, age, sex, and race

Chapter III: Low-Concentration PM$_{2.5}$ and Mortality: Estimating Acute and Chronic Effects in a Population Based Study

Figure 1A Mean PM$_{2.5}$ concentrations in 2004 at a high resolution (1 × 1 km) across New England predicted by the AOD models

Figure 1B Predicted PM$_{2.5}$ concentrations at a 1 × 1 km grid for November 15, 2003

Figure 2 Percent change in mortality per 10 µg/m$^3$ increase in short-term PM$_{2.5}$ with different lags with mutual adjustment. Error bars indicate the 95% Cis

Figure 3 The dose-response relationship between long-term PM$_{2.5}$ and mortality at low doses with mutual adjustment (a) and the dose-response relationship between short-term PM$_{2.5}$ and mortality at low doses without mutual adjustment (b). Shaded areas indicate the 95% Cis
List of Tables with Captions

Chapter I: Estimating daily air temperature across the Southeastern United States using high-resolution satellite data: a statistical modeling study

Table 1 Prediction accuracy in Stage 1: 10-fold cross-validated (CV) $R^2$ for daily $T_a$ predictions (2000-2014) .................................................................15

Table 2 Prediction accuracy in Stage 3: $R^2$ for daily $T_a$ predictions (2000-2014) ........18

Table 3 Contribution of each stage for the daily $T_a$ predictions for each year (2000-2014) .........................................................................................18

Table 4 Prediction accuracy: $R^2$ for comparing final daily $T_a$ predictions with MERRA products (2000-2014) .........................................................22

Chapter II: Impacts of Temperature and its Variability on Mortality in New England

Table S1 Characteristics of the study population across New England for the years 2000-2008 .................................................................45

Table S2 Descriptive Statistics for temperature and temperature variability in summer and winter across New England, 2000-2008 ........................................46

Table S3 Correlation coefficients ($r$) of the spatial and temporal variations for mean temperature and temperature variability .............................................47
Table S4 Percent increase in mortality (95% CI) for each 1 °C increase for both seasonal mean temperature and temperature SD .................................................48

Chapter III: Low-Concentration PM$_{2.5}$ and Mortality: Estimating Acute and Chronic Effects in a Population Based Study

Table 1 Descriptive Statistics for PM$_{2.5}$ exposure and temperature in New England, 2003-2008 .................................................................66

Table 2 Percent increase in mortality (95% CI) for a 10 µg/m$^3$ increase for both short-term and long-term PM$_{2.5}$ .........................................................67
Acknowledgements

I would first like to express my deepest gratitude to my advisor, Professor Joel Schwartz, for his encouragement, support, and guidance throughout the years. He gave me tremendous supports and freedom to try my own ideas and insightful suggestions for data analysis and interpretation. It is my great honor to join his lab and work with him.

I have been extremely fortunate to pursue my ScD at Harvard University, with exposure to so many great scientists. I thank the members of my thesis committee and qualifying exam committee: Petros Koutrakis, Brent Coull, and Francine Laden for their insightful comments and supports. All of them were eager to help me and connect me with resources that would strengthen my research. I also gratefully acknowledge my co-authors, lab mates, classmates, and the faculty and staff at HSPH for their support on this journey. I would like to thank the USEPA and NIEHS grants for funding my research.

Lastly, I dedicate this dissertation to my wonderful, loving family. I thank my mother and father, and my little sister, for their unconditional love and support of all my endeavors. Most of all, I thank my husband, Pengfei Liu, for encouraging me to follow my dreams and for accompanying me through the wonderful four years at Harvard.

Liuhua Shi
Previously Published Work

Results in Chapter 1 were published as


Results in Chapter 2 were published as


Results in Chapter 3 were published as

CHAPTER I

Estimating daily air temperature across the Southeastern United States using high-resolution satellite data: a statistical modeling study

Liuhua Shi 1, Pengfei Liu 2, Itai Kloog 3, Mihye Lee 1, Anna Kosheleva 1, Joel Schwartz 1

1Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA

2John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

3Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Beer Sheva, Israel
ABSTRACT

Accurate estimates of spatio-temporal resolved near-surface air temperature ($T_a$) are crucial for environmental epidemiological studies. However, values of $T_a$ are conventionally obtained from weather stations, which have limited spatial coverage. Satellite surface temperature ($T_s$) measurements offer the possibility of local exposure estimates across large domains. The Southeastern United States has different climatic conditions, more small water bodies and wetlands, and greater humidity in contrast to other regions, which add to the challenge of modeling air temperature. In this study, we incorporated satellite $T_s$ to estimate high resolution (1 km $\times$ 1 km) daily $T_a$ across the southeastern USA for 2000-2014. We calibrated $T_s$ to $T_a$ measurements using mixed linear models, land use, and separate slopes for each day. A high out-of-sample cross-validated $R^2$ of 0.952 indicated excellent model performance. When satellite $T_s$ were unavailable, linear regression on nearby monitors and spatio-temporal smoothing was used to estimate $T_a$. The daily $T_a$ estimations were compared to the NASA’s Modern-Era Retrospective Analysis for Research and Applications (MERRA) model. A good agreement with an $R^2$ of 0.969 and a mean squared prediction error (RMSPE) of 1.376 °C was achieved. Our results demonstrate that $T_a$ can be reliably predicted using this $T_s$-based prediction model, even in a large geographical area with topography and weather patterns varying considerably.
INTRODUCTION

Global warming adds urgency to better understand the impact of temperature on health, particularly in the warm areas such as the southeastern USA. Growing evidence has linked near-surface air temperature ($T_a$), an important environmental stressor, with morbidity and mortality (1-7). Previous studies concerning human health and $T_a$ exposure are primarily limited by the spatial and temporal availability of $T_a$ measurements, leaving large areas uncovered. Temperature can vary greatly both in space and time, therefore these collected point-samples are insufficient to adequately capture the spatial and temporal variability within a large area (8). In addition, owing to the urban heat-island effect, higher temperatures are often observed in urban areas versus surrounding areas. For example, temperatures measured in a station near an airport may underestimate the true near-surface air temperatures in the urban area.

However, these small geographic changes in air temperature can have important health effects. Shi et al. (2015) found that both geographical contrasts and annual anomalies of local air temperature contribute to excess public health burden of climate change (9). They used daily local air temperature estimates at fine geographic scale (1 km × 1 km) in New England, to capture the exposure variability in space and time which may be driving the adverse health effects. Lower exposure measurement errors, and the inclusion of the entire region, largely alleviate the downward bias in health effect estimates.

In recent years, several methods were developed to address the lack of high-resolution exposure data. Geospatial statistical methods, such as land use regression and kriging, are most commonly used approaches. They allow characterizing the spatial
heterogeneity of exposure by using time invariant geographical variables to expand the ground monitored measurements to large areas (10, 11). However these methods do not generally capture temporal variability in exposure, in that they are commonly based on a year of intensive monitoring, and miss changes over years in the spatial pattern of $T_a$. Therefore they are primarily used to assess chronic health effects.

Satellite-based remote sensing can provide additional information at high spatial and temporal resolution. Satellite instruments, such as Moderate Resolution Imaging Spectroradiometer (MODIS), can provide global daily estimates of 1 km surface skin temperature ($T_s$), the temperature at the air-soil interface (12-14). $T_s$ is derived from the thermal infrared signal received by the satellite sensor. $T_s$ as an indicator of the net surface energy balance, depends on the presence of vegetation or plant cover, atmospheric conditions, and thermal properties of the underlying surfaces. It is different from $T_a$, which is measured at meteorological stations at the screen height of 1-2 m above the land surface.

Several studies have shown that $T_s$ and $T_a$ are correlated (15, 16). Even so, there are many factors that can influence the complex and geographically heterogeneous relationship between $T_s$ and $T_a$, such as humidity, the type of underlying surfaces, the elevation, and other surface parameters. Additionally, Dousset (17) stated that $T_a$ has superior correlation with $T_s$ at night because the solar radiation does not affect the thermal infrared signal. During nighttime $T_s$ is close to $T_a$ and during daytime $T_s$ is generally higher than $T_a$ (12).

The value of $T_a$ cannot be predicted by $T_s$ using a simple linear relationship with reasonable accuracy. Fu et al. (2011) explored predicting $T_a$ using satellite $T_s$ and found
an $R^2 > 0.55$ (18). Recently, Kloog et al. (2014) presented a novel model and assessed daily mean $T_a$ in the northeastern USA using MODIS-derived $T_s$ measurements (19, 20). Better predictive performance was reported. For days with available $T_s$ data, mean out-of-sample $R^2$ was 0.947. Even for days without $T_s$ values, the model accuracy was also excellent. Although $T_a$ estimation in the northeastern USA was excellent, it is uncertain that how well the satellite approach would perform in areas with different geographic features and weather patterns.

Predicting $T_a$ for the southeastern USA is of top priority for epidemiological studies. Many researchers are of great interest in investigating the health effects of temperature using the Medicare cohort (aged 65+), because this elderly population is potentially most vulnerable to climate change. Due to its warm weather, the southeastern USA contains a very large Medicare population (13 millions). Thus, it is particularly urgent to provide a comprehensive temperature dataset for this particular region.

The aim of this paper was to estimate 15-year daily local air temperature in the southeastern USA, by extending and validating the previous hybrid-model approach to account for the unique geographical and climatological characteristics of the study area. Specifically, by incorporating satellite derived $T_s$, $T_a$ measured at monitors, meteorological variables and land use terms, we employed a 3-stage statistical modeling approach to obtain daily air temperature predictions at 1 km × 1 km resolution across the southeastern USA for the years 2000-2014. The retrieved air temperature was cross validated against the measurements from weather stations. As an independent validation, the results were also compared with NASA’s reanalysis products, Modern-Era Retrospective Analysis for Research and Applications (MERRA).
MATERIALS AND METHODS

Study domain

The spatial domain of our study includes the southeastern part of the USA, comprising the states of Georgia, Alabama, Mississippi, Tennessee, North Carolina, South Carolina and Florida (Fig. 1). The southeastern states include some populous metropolitan areas (Charlotte, Memphis, Raleigh, Atlanta and Miami), rural towns, large forested regions, mountains, water bodies, and the Atlantic sea shoreline. The study region covers an area of 916,904 km$^2$ with a population of 56,742,948 according to the 2010 census (UCSB, 2010), and encompasses 1,013,408 discrete 1 km $\times$ 1 km satellite grid cells.

Fig. 1 Map of the study area showing all available NCDC air temperature monitor stations across southeastern USA for 2000-2014
Surface temperature

Daily $T_s$ data from the MODIS sensors located on polar orbiting Terra satellite for the years 2000-2014, with a spatial resolution of 1 km x 1 km were used to prepare the analysis presented in this paper. They were calculated from land surface temperature (LST) and emissivity using the formula $T_s = \frac{\text{LST}}{\text{Emissivity}^{1/4}}$. These MOD11_A1 products (LST & emissivity) are freely available online through the NASA website (NASA, 2014), and cover tiles h10v05, h10v06, and h11v05. They include the latitude and longitude of 1 km grid cell centroid, nighttime and daytime LST, and emissivity.

Nighttime $T_s$ was observed to have a higher correlation with air temperature ($r = 0.94$), compared to daytime $T_s$ ($r = 0.87$). This is consistent with previous literature, demonstrating that the nighttime $T_s$ generally present superior correlations with ground measurements of $T_a$ across most sites in the USA. More details about MODIS $T_s$ data can be found in the literature (19, 21).

Meteorological data from weather stations

In our analysis, daily data for $T_a$ from weather stations across the southeastern USA (see Fig. 1) for 2000-2014 were obtained from the National Climatic Data Center (NCDC). There were 538 NCDC stations operating daily in the southeastern USA during the study period. The mean $T_a$ across the study area during the study period was 18.59 ºC with a standard deviation of 7.96 ºC and an interquartile range (IQR) of 12.28 ºC.
NDVI

The presence of vegetation on the surface can also affect $T_s$, since incoming solar radiation is partly intercepted by vegetative surfaces during the day, and part of the outgoing longwave radiation at night is also intercepted by vegetation. Normalized Difference Vegetation Index (NDVI) data were available at 1 km resolution from the monthly MODIS product (MOD13A3). Monthly vegetation index was used as NDVI values have little within-month variation and the spatial distribution is illustrated in the Appendix A (Fig. A.1). To create the covariate of NDVI for daily $T_s$, the distances between centroids of NDVI and that of a grid cell of $T_s$ were calculated, and the nearest NDVI value of the current month was merged.

Spatial predictors of air temperature

To enhance the predictive ability of the final model, we included the following statistically significant spatial predictors in the models, which influence $T_s$ or $T_a$ and thereby modify the relationship between $T_s$ and $T_a$. These time-invariant variables were first processed into 1 km × 1 km spatial resolution (details are described below). Spatial distributions for these variables are illustrated in the Appendix A (Fig. A.2 – Fig. A.4). For the daily dataset, spatial predictors were assigned to the nearest grid cell for each day, and these data served as covariates for the daily surface temperature.

Percent of urban areas

In urban areas, residential, commercial and industrial developments often produce radical changes in radiative, thermodynamic and aerodynamic characteristics of the surface. Thus the associated $T_s$ are often modified substantially. Percent of urban areas data were
obtained from the 2011 national land cover data (NLCD) (22). Data were available as raster files with a 30 m spatial resolution. We reclassified the raster into 0 (open space) and 1 (urban areas), by recoding land cover codes 22, 23, 24 (sub categories for developed areas) to 1 as an urban cell and assigning 0 for the remaining. The mean of the 30 m-resolution binary values within each 1 km grid cell was calculated, namely percent of urban areas.

**Elevation**

There are notable elevation contrasts across such a large area, thus elevation was used as a spatial predictor. Generally, higher elevations are associated with lower air temperatures. Elevation data were obtained from the National Elevation Dataset (23). NED data is available from the U.S. Geological Survey (USGS) and provides elevation data covering the United States at a spatial resolution of 1 arc second (30 m). The mean of the 30 m-resolution elevation values within each 1 km grid cell was calculated and used as a spatial predictor.

**Distance to water body**

The presence of moisture at the surface greatly moderates the diurnal range of surface temperatures, due to the increased evaporation from the surface and increased heat capacity of water. Over a free water body, about 80% of the net radiation is utilized for evaporation on average and the ground heat flux is reduced. We used the Esri Data and Maps for ArcGIS 2013 for water body data. We created a dummy variable taking the value of 1, if the 1 km grid cell centroid is on water and 0 otherwise. The distance from centroids to water body (a continuous measure) was also calculated, which equals to 0 if the centroid is on water.
STATISTICAL METHODS

Data preparation was implemented using MATLAB (R2014b, MathWorks), and all modeling was done using the R statistical software. The prediction process consists of 3 stages. The stage 1 model calibrates the $T_s - T_a$ relationship on each day during 2000-2014 using data from grid cells with both ground $T_a$ monitors and $T_s$ measurements. We performed daily calibrations with nighttime MODIS $T_s$ for the reasons noted earlier. The base model (stage 1), fit to data from each year (2000-2014) separately, consists of a mixed model with day-specific random effects to capture the day-to-day variation in the $T_s - T_a$ relationship. Specifically the base model can be written as:

$$T_{a_{ij}} = (\alpha + u_{j}) + (\beta_1 + v_j) T_{s_{ij}} + \beta_2 \text{Elevation}_i + \beta_3 \text{NDVI}_{ik} + \beta_4 \text{Percent urban}_i + \beta_5 \text{Distance to water body}_i + \beta_6 \text{water body}_i + \epsilon_{ij}$$

$$(u_j, v_j) \sim N([0 0], \Sigma)$$

where $T_{a_{ij}}$ denotes the measured mean air temperature at a spatial site $i$ on a day $j$; $\alpha$ and $u_j$ are the fixed and random (day-specific) intercepts, respectively, $T_{s_{ij}}$ is the surface temperature value in the grid cell corresponding to site $i$ on a day $j$; $\beta_1$ and $v_j$ are the fixed and random slopes, respectively. $\text{Elevation}_i$ is the mean elevation in the grid cell corresponding to site $i$, $\text{NDVI}_{ik}$ is the monthly NDVI value in the grid cell corresponding to site $i$ for the month in which day $j$ falls, percent urban$_i$ is the percent of urban area in the grid cell, distance to water body$_i$ is the distance of the grid cell to the nearest water body, and water body$_i$ is a 0/1 dummy variable identifying grid cells that intersect water polygons. Finally, $\Sigma$ is an unstructured variance-covariance matrix for the random effects and $\epsilon_{ij}$ is the error term at site $i$ on a day $j$. 
The performance of the base model was validated by 10-fold cross-validation. The dataset was randomly divided into 90% and 10% splits ten times. We fit the model using 90% of the data, and then use this model to predict for the held-out 10% of the data. Then the “out-of-sample” cross-validated (CV) \( R^2 \) were computed. To check for bias, we regressed the measured \( T_a \) values in the held-out data against the predicted values in each site on each day. We assessed the model prediction performance by taking the square root of the mean squared prediction errors (RMSPE). Overall temporal \( R^2 \) and overall spatial \( R^2 \) were calculated as well. More details are provided in Kloog et al. (20).

In stage 2, we predicted \( T_a \) in grid cells without monitors but with available \( T_s \) measurements using the stage 1 model coefficients. This is implemented by simply applying the estimated prediction model fit obtained from stage 1 to these additional \( T_s \) values. This resulted in datasets with \( T_a \) predictions for all available \( T_s \) cell/day combinations.

To impute data for grid cells/days for which \( T_s \) measurements were not available, the stage 3 model was fitted by using the stage 2 predictions. Specifically, for each grid cell, we regressed the \( T_a \) predictions obtained from stage 2 on the daily mean measured \( T_a \) (from the stations within a 100 km buffer of that grid cell), land use terms and a smooth nonparametric function of latitude and longitude of the grid cell centroid, with random cell-specific intercepts and slopes. This is similar to universal kriging, by using \( T_a \) measurements from nearby grid cells to help fill in the missing. We selected a 100 km buffer size because it was small enough to ensure relevance and large enough to include multiple \( T_a \) stations to produce more stable estimates. Because the spatial patterns of \( T_a \) vary temporally, a separate spatial surface was fit for each two-month period.
approach provides additional information about the $T_a$ in the missing grid cells that simple kriging would not. Specifically, the following semiparametric regression model was fitted:

$$\text{Pred}T_{aij} = (a + u_j) + (\beta_1 + v_i) mT_{aij} + \beta_2 \text{NDVI}_{ik} + \beta_3 \text{Percent urban}_i + \beta_4 \text{Distance to water body}_i + \text{Smooth} (X,Y) k(j) + \text{Bimon} + \varepsilon_{ij}$$

$$(u_j, v_i) \sim [(0, 0), \Omega_{ij}]$$

where $\text{Pred}T_{aij}$ is the predicted air temperature at a grid cell $i$ on a day $j$ from the mixed model; $mT_{aij}$ is the mean $T_a$ in the relevant 100 km buffer for cell $i$ on a day $j$; $a$ and $u_j$ are the fixed and random intercepts, respectively; $\beta_1$ and $v_i$ are the fixed and random slopes, respectively. NDVI$_{ik}$ is the monthly NDVI value in the grid cell corresponding to site $i$ for the month in which day $j$ falls, percent urban$_i$ is the percent of urban area in the grid cell, and distance to water body$_i$ is the distance of the grid cell to nearest water body. The smooth $(X,Y)$ is a thin plate spline fit to the latitude and longitude of the centroid of grid cell $i$, $k(j)$ denotes the two-month period in which day $j$ falls (that is, a separate spatial smooth was fit for each two-month period).

The calculated coefficients of the stage 3 model would then be used to fill in the missing $T_a$ values. In contrast to stage 2, the stage 3 model includes cell-specific random intercepts and slopes, which allows for temporal and spatial interpolation for each grid cell. That is, we could use the random effects for a grid cell to help impute $T_a$ data of this cell for days when $T_s$ measurements were unavailable. If grid cells did not have any temperature monitors within their 100 km buffer, no temperature was imputed. Such locations were generally in unpopulated areas.
Validation against reanalysis data

MERRA is a NASA reanalysis dataset generated using the version 5.2.0 of the Goddard Earth Observing System (GEOS-5) atmospheric data assimilation system. Reanalysis is a technique for generating a comprehensive meteorological record by assimilating observations from multiple platforms using numerical models. Different from the statistical approach used in this study, the models used in reanalysis are based on physical rules of atmospheric motions. The reanalysis datasets are widely used for analysis of long term, large scale climatic changes. Their spatial resolution, however, is usually coarse (24). Although reanalysis may also utilize some of the observations, such as the weather station data and satellite irradiance, which are also used in the statistical modeling, it represents an entirely different modeling approach and its results can be considered as an independent source of information for validation purpose.

In this study, the MERRA daily air temperature at 2 m above displacement height ($T_{2m}$) was chosen to validate the retrieved $T_a$. MERRA data has a coarser spatial resolution of $1/2^\circ \times 2/3^\circ$ in latitude/longitude. The 1 km grids of retrieved $T_a$ were matched with the nearest MERRA grid centroid. A MERRA grid can contain 3000 to 4300 of such 1 km grids. These grids were aggregated and averaged by day, and the averaged data were used for validation. For some MERRA grids which contain both land and ocean pixels, the number of matched 1 km grids can be less, because our statistical model only retrieves air temperature above the land. To avoid interference of ocean pixels, MERRA with less than 1878 (lowest 15th percentile) 1 km grids matched were removed from analysis.
RESULTS

The correlation between the satellite-derived $T_s$ and the $T_a$ obtained from the monitor sites within the same grid cells was high, indicating that $T_s$ can provide useful information for predicting $T_a$. As an example, results for the year 2011 are shown in Fig. 2 (left). An $R^2$ value of 0.89 indicates that although $T_s$ and $T_a$ are highly correlated, there are still some variations of $T_a$ cannot be explained by $T_s$ without the calibration using more predictors. Other factors, such as elevation, NDVI, and land use terms can also affect the $T_a$ - $T_s$ relationship. Taking those factors into account, the relationship between the predicted $T_a$ from our daily calibration method and the monitored $T_a$ is greatly improved (e.g., $R^2$=0.96 in 2011).

![Fig. 2](image)

**Fig. 2** Scatter plots of the monitored air temperature versus satellite surface temperature (left) and the monitored air temperature versus that from the stage 1 calibration (right). Data are shown for the year 2011.
For the stage 1 calibration, we conducted a 10-fold cross-validation. The results of 10-fold cross-validation, which can represent the out-of-sample performance of our stage 1 model, are presented in Table 1. For the entire study period, the mean out-of-sample $R^2$ was 0.952 (year-to-year variation 0.935 - 0.962). The spatial and temporal $R^2$ values (mean of 0.867 and 0.961, respectively) also showed good performance of the model. We found no bias in our cross-validation results (slope of observed versus predicted = 1.00). The model also yielded small prediction errors, with an RMSPE of 1.662 °C and spatial RMSPE of 0.903 °C. All of these indicated excellent model performance.

**Table 1**
Prediction accuracy in Stage 1: 10-fold cross-validated (CV) $R^2$ for daily $T_a$ predictions (2000-2014)

<table>
<thead>
<tr>
<th>Year</th>
<th>CV $R^2$</th>
<th>CV $R^2_{\text{Spatial}}$</th>
<th>CV $R^2_{\text{Temporal}}$</th>
<th>RMSPE</th>
<th>RMSPE$_{\text{Spatial}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>0.941</td>
<td>0.804</td>
<td>0.955</td>
<td>1.74</td>
<td>1.071</td>
</tr>
<tr>
<td>2001</td>
<td>0.937</td>
<td>0.860</td>
<td>0.948</td>
<td>1.796</td>
<td>0.973</td>
</tr>
<tr>
<td>2002</td>
<td>0.962</td>
<td>0.888</td>
<td>0.970</td>
<td>1.530</td>
<td>0.871</td>
</tr>
<tr>
<td>2003</td>
<td>0.944</td>
<td>0.840</td>
<td>0.955</td>
<td>1.753</td>
<td>0.942</td>
</tr>
<tr>
<td>2004</td>
<td>0.943</td>
<td>0.858</td>
<td>0.953</td>
<td>1.804</td>
<td>0.938</td>
</tr>
<tr>
<td>2005</td>
<td>0.960</td>
<td>0.854</td>
<td>0.971</td>
<td>1.541</td>
<td>0.888</td>
</tr>
<tr>
<td>2006</td>
<td>0.935</td>
<td>0.792</td>
<td>0.942</td>
<td>1.920</td>
<td>0.977</td>
</tr>
<tr>
<td>2007</td>
<td>0.957</td>
<td>0.834</td>
<td>0.967</td>
<td>1.544</td>
<td>0.924</td>
</tr>
<tr>
<td>2008</td>
<td>0.950</td>
<td>0.891</td>
<td>0.958</td>
<td>1.738</td>
<td>0.848</td>
</tr>
<tr>
<td>2009</td>
<td>0.950</td>
<td>0.875</td>
<td>0.958</td>
<td>1.812</td>
<td>0.897</td>
</tr>
<tr>
<td>2010</td>
<td>0.960</td>
<td>0.917</td>
<td>0.969</td>
<td>1.709</td>
<td>0.871</td>
</tr>
<tr>
<td>2011</td>
<td>0.962</td>
<td>0.901</td>
<td>0.970</td>
<td>1.487</td>
<td>0.878</td>
</tr>
<tr>
<td>2012</td>
<td>0.959</td>
<td>0.890</td>
<td>0.967</td>
<td>1.458</td>
<td>0.810</td>
</tr>
<tr>
<td>2013</td>
<td>0.960</td>
<td>0.919</td>
<td>0.965</td>
<td>1.562</td>
<td>0.826</td>
</tr>
<tr>
<td>Year</td>
<td>2014</td>
<td>2015</td>
<td>2016</td>
<td>2017</td>
<td>2018</td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Overall Mean</td>
<td>0.952</td>
<td>0.867</td>
<td>0.961</td>
<td>1.662</td>
<td>0.903</td>
</tr>
</tbody>
</table>

In stage 2, we used the parameters derived from stage 1 model to predict $T_a$ in grid cells without monitors but with available $T_s$ measurements. Fig. 3 (top) presents the spatial pattern of predicted $T_a$ values from stage 2 at $1 \text{ km} \times 1 \text{ km}$ resolution, on a sample day (2011.08.25). It can be seen that there are a considerable number of grid cells with missing values, because of the missing satellite data, typically because of cloud cover. For the entire 15 years, stage 2 contributed to 42.4% of the final $T_a$ predictions.
Fig. 3 Predicted air temperature (°C) from stage 2 (top) and both stage 2 & 3 (bottom) in each 1 km × 1 km grid on a sample day (2011.08.25) across the southeastern USA.

The missing values of stage 2 can be filled using a stage 3 model, and the results are shown in Fig. 3 (bottom). The stage 3 model also performed well. The $R^2$ values between stage 3 and stage 2 predictions are shown in Table 2, and the mean $R^2$ value for the entire
study period is 0.971 (year-to-year variation 0.962 - 0.981). For the entire 15 years, 56.2% of the final predictions were derived from stage 3. There was still a small amount of missing predictions (1.4%) due to lack of monitored $T_a$ measurements within 100 km buffer on a day. More details are in Table 3.

Table 2. Prediction accuracy in Stage 3: $R^2$ for daily $T_a$ predictions (2000-2014)

<table>
<thead>
<tr>
<th>Year</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>0.962</td>
</tr>
<tr>
<td>2001</td>
<td>0.964</td>
</tr>
<tr>
<td>2002</td>
<td>0.972</td>
</tr>
<tr>
<td>2003</td>
<td>0.970</td>
</tr>
<tr>
<td>2004</td>
<td>0.969</td>
</tr>
<tr>
<td>2005</td>
<td>0.973</td>
</tr>
<tr>
<td>2006</td>
<td>0.969</td>
</tr>
<tr>
<td>2007</td>
<td>0.972</td>
</tr>
<tr>
<td>2008</td>
<td>0.973</td>
</tr>
<tr>
<td>2009</td>
<td>0.976</td>
</tr>
<tr>
<td>2010</td>
<td>0.981</td>
</tr>
<tr>
<td>2011</td>
<td>0.972</td>
</tr>
<tr>
<td>2012</td>
<td>0.969</td>
</tr>
<tr>
<td>2013</td>
<td>0.974</td>
</tr>
<tr>
<td>2014</td>
<td>0.975</td>
</tr>
<tr>
<td>Overall Mean</td>
<td>0.971</td>
</tr>
</tbody>
</table>

Table 3

Contribution of each stage for the daily $T_a$ predictions for each year (2000-2014)

<table>
<thead>
<tr>
<th>Year</th>
<th>Stage 2</th>
<th>Stage 3</th>
<th>Missing</th>
</tr>
</thead>
</table>

18
<table>
<thead>
<tr>
<th>Year</th>
<th>Overall Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>38.1% 60.4% 1.5%</td>
</tr>
<tr>
<td>2001</td>
<td>39.4% 59.2% 1.4%</td>
</tr>
<tr>
<td>2002</td>
<td>37.7% 60.9% 1.4%</td>
</tr>
<tr>
<td>2003</td>
<td>39.4% 59.1% 1.5%</td>
</tr>
<tr>
<td>2004</td>
<td>38.2% 60.6% 1.2%</td>
</tr>
<tr>
<td>2005</td>
<td>44.3% 54.6% 1.1%</td>
</tr>
<tr>
<td>2006</td>
<td>47.2% 51.7% 1.1%</td>
</tr>
<tr>
<td>2007</td>
<td>45.7% 53.2% 1.1%</td>
</tr>
<tr>
<td>2008</td>
<td>46.3% 52.6% 1.1%</td>
</tr>
<tr>
<td>2009</td>
<td>41.0% 57.9% 1.1%</td>
</tr>
<tr>
<td>2010</td>
<td>46.2% 52.2% 1.6%</td>
</tr>
<tr>
<td>2011</td>
<td>48.0% 50.0% 2.0%</td>
</tr>
<tr>
<td>2012</td>
<td>45.2% 52.9% 1.9%</td>
</tr>
<tr>
<td>2013</td>
<td>39.0% 59.7% 1.3%</td>
</tr>
<tr>
<td>2014</td>
<td>40.3% 58.6% 1.1%</td>
</tr>
<tr>
<td>Overall Mean</td>
<td>42.4% 56.2% 1.4%</td>
</tr>
</tbody>
</table>

Figure 4 illustrates the spatial pattern of estimated $T_a$ values from the 3-stage statistical models, averaged over the year of 2011 for the southeastern USA area. In the calculation of the annual average, we removed grid cells with less than 243 observations (2/3 of the daily $T_a$ predictions) in 2011. After restriction, estimated annual average $T_a$ values across the study area for 2011 ranged from 10.91 °C to 25.91 °C. The figure shows that urban areas such as Raleigh, Atlanta, Birmingham, Jackson, Columbia, and Memphis, appear warmer than the surrounding areas possibly due to urban heat-island effect (25). In addition, estimated annual $T_a$ values are higher in areas closer to the shoreline compared with inland areas as an impact of the ocean.
Fig. 4 Spatial pattern of predicted air temperature (°C), averaged over the 2011 for the southeastern USA.

In addition to the spatial variability we observed, air temperature showed temporal variation in its spatial pattern as well. Figure 5 shows that there is space-time variation in the estimated $T_a$ values for a subset of the study area. The results indicate that our model predictions can well resolve the day-to-day variation in the spatial pattern of $T_a$ exposure.
Fig. 5 Predicted air temperature (°C) at a 1 km × 1 km grid for August 3, 2011 (top) and August 25, 2011 (bottom).

With respect to the final $T_a$ exposure dataset that we generated, an independent validation was performed, by comparing our final daily $T_a$ dataset with MERRA products $T_{2m}$. A good agreement was achieved, with an overall $R^2$ of 0.969 and RMSPE of 1.376 °C. The spatial and temporal $R^2$ values are 0.954 and 0.976, respectively. Results for the
year 2011 are shown in Fig. 6 as an example. Comparison results for 15 years are compiled in Table 4.

![Fig. 6 A scatter plot of the calculated daily average $T_a$ versus the daily $T_{2m}$ from MERRA. Data are shown for the year 2011.](image)

Table 4
Prediction accuracy: $R^2$ for comparing final daily $T_a$ predictions with MERRA products (2000-2014)

<table>
<thead>
<tr>
<th>Year</th>
<th>CV $R^2$</th>
<th>CV $R^2_{\text{Spatial}}$</th>
<th>CV $R^2_{\text{Temporal}}$</th>
<th>RMSPE</th>
<th>RMSPE$_{\text{Spatial}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>0.959</td>
<td>0.896</td>
<td>0.970</td>
<td>1.553</td>
<td>1.375</td>
</tr>
<tr>
<td>2001</td>
<td>0.964</td>
<td>0.955</td>
<td>0.972</td>
<td>1.388</td>
<td>1.095</td>
</tr>
<tr>
<td>2002</td>
<td>0.972</td>
<td>0.966</td>
<td>0.978</td>
<td>1.344</td>
<td>0.846</td>
</tr>
<tr>
<td>2003</td>
<td>0.974</td>
<td>0.958</td>
<td>0.981</td>
<td>1.245</td>
<td>0.876</td>
</tr>
<tr>
<td>2004</td>
<td>0.969</td>
<td>0.956</td>
<td>0.978</td>
<td>1.356</td>
<td>0.979</td>
</tr>
<tr>
<td>2005</td>
<td>0.971</td>
<td>0.955</td>
<td>0.979</td>
<td>1.328</td>
<td>0.871</td>
</tr>
<tr>
<td>2006</td>
<td>0.960</td>
<td>0.952</td>
<td>0.970</td>
<td>1.487</td>
<td>1.261</td>
</tr>
<tr>
<td>Year</td>
<td>2007</td>
<td>2008</td>
<td>2009</td>
<td>2010</td>
<td>2011</td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Ta 1</td>
<td>0.965</td>
<td>0.968</td>
<td>0.971</td>
<td>0.978</td>
<td>0.969</td>
</tr>
<tr>
<td>Ta 2</td>
<td>0.951</td>
<td>0.964</td>
<td>0.965</td>
<td>0.953</td>
<td>0.964</td>
</tr>
<tr>
<td>Ta 3</td>
<td>0.971</td>
<td>0.974</td>
<td>0.981</td>
<td>0.984</td>
<td>0.976</td>
</tr>
<tr>
<td>Ta 4</td>
<td>1.456</td>
<td>1.389</td>
<td>1.342</td>
<td>1.361</td>
<td>1.423</td>
</tr>
<tr>
<td>Ta 5</td>
<td>1.154</td>
<td>0.942</td>
<td>0.970</td>
<td>0.840</td>
<td>0.969</td>
</tr>
</tbody>
</table>

**DISCUSSION**

The southeastern USA is distinguished by its warm weather and serves as a place for relocation, particularly for the elderly who are most sensitive to climate change. The rapid population growth further urges epidemiological studies on the health effects of climate change in this area. Most temperature-related epidemiological studies have relied on central meteorological stations to assess $T_a$. The poor spatial resolution of monitor stations can result in exposure measurement error, with expected downward bias in estimating the health effects of temperature and elevated risk of not finding a significant association. This has been recently demonstrated in a study of temperature and birth weight in Massachusetts (27). As a result, recently increasing epidemiological studies have used the finer scale exposure estimates from remote sensing (28-30).

To the best of our knowledge, this work is the first study estimating spatially and temporally resolved air temperature for the southeastern USA with a humid subtropical climate. We presented several key features in our study. First, the $T_a$ prediction models
exhibited excellent model performances. We had no bias in the cross-validation results (average slope of 1.00). In contrast to the previous work by Kloog et al (2014) in the northeastern USA, we fit a more complex model for stage 3, which included important land use terms such as proximity to water, urbanization, and time varying covariates such as NDVI. Consequently, our stage 3 model had a higher $R^2$ (0.971) than was observed in the previous work (0.940). The results also compared well with reanalysis products, which were generated using an entirely different approach from statistical modeling.

Another key feature is that this study provides local temperature estimates on a daily basis within a study area for 15 years. Such a large 15-year daily exposure dataset will be particularly valuable for estimating the health effects of temperature in both short- and long-term, and locations outside of metropolitan areas, which are rarely included in epidemiology studies of temperature. In sum, our analysis showed that regardless of the quantity or quality of predictions, $T_a$ can be reliably predicted from $T_s$ if modeled appropriately.

Despite the high accuracy and low percent of missing of our model prediction, there are several limitations in the present study. For one the southeastern USA has many wetlands (swamps), which is part of the challenge of retrieving satellite $T_s$ information. Even though 1 km is the finest resolution of $T_s$ products that are available, Sabrino and coauthors reported that spatial resolutions lower than 50 m would underestimate the urban heat-island effect (31). In addition, this satellite-based approach still depends on ground $T_a$ monitors. Without sufficient $T_a$ monitors, the predictive power might be reduced dramatically, especially the mixed linear model. Because the mixed linear model uses a daily calibration method, thus requires relatively densely distributed daily $T_a$
stations. In addition to ground $T_a$ monitors, this satellite-based approach is also affected by some other factors, such as atmospheric conditions. Furthermore, daytime and nighttime temperatures detected by satellite are not daily mean temperatures. As a caveat, the $T_a$ retrieved from this study should be considered as a high-resolution alternative for those areas with limited meteorological stations but cannot replace ground monitoring temperature.

CONCLUSION

In summary, we demonstrate how satellite-derived surface temperature can be used reliably to construct a high resolution dataset of daily air temperature. This dataset could well resolve and capture the spatial and temporal variations in $T_a$ in warm areas. Public health burden of these variations should be investigated in future studies.
REFERENCES


CHAPTER II

Impacts of Temperature and its Variability on Mortality in New England

Liuhua Shi\textsuperscript{1}, Itai Kloog\textsuperscript{2}, Antonella Zanobetti\textsuperscript{1}, Pengfei Liu\textsuperscript{3}, Joel D. Schwartz\textsuperscript{1,4}

\textsuperscript{1}Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

\textsuperscript{2}Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Beer Sheva, Israel

\textsuperscript{3}School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA

\textsuperscript{4}Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
Rapid buildup of greenhouse gases is expected to increase the Earth surface mean temperature, with unclear effects on temperature variability\textsuperscript{1-3}. This adds urgency to better understand the direct effects of the changing climate on human health. However, the effects of prolonged exposures to temperatures, which are important for understanding the public health burden, are unclear. Here we demonstrate that long-term survival was significantly associated with both seasonal mean values and standard deviations (SDs) of temperature among the Medicare population (aged 65+) in New England, and break that down into long-term contrasts between ZIP codes and annual anomalies. A rise in summer mean temperature of 1 °C was associated with 1.0% higher death rate whereas an increase in winter mean temperature corresponded to 0.6% lower mortality. Increases in temperature SDs for both summer and winter were harmful. The increased mortality in warmer summers was entirely due to anomalies, while it was long term average differences in summer SD across ZIP codes that drove the increased risk. For future climate scenarios, seasonal mean temperatures may in part account for the public health burden, but excess public health risk of climate change may also stem from changes of within season temperature variability.

Emissions of greenhouse gases will change the Earth’s climate, most notably by changing temperature and temperature variability\textsuperscript{1}. The Intergovernmental Panel on Climate Change (IPCC), in its newly released fifth assessment report, forecasts a rise in world average temperature ranging from 0.2-5.5 °C by 2100\textsuperscript{1}. This has increased interest in the impact of temperature on health.
While many studies have reported associations between short-term temperature changes and increased daily deaths\textsuperscript{4-11}, evidence on the association between annual mortality and changes in seasonal temperature averages is scarce\textsuperscript{12,13}. Since such impacts by definition cannot be short-term mortality displacement (we are looking at annual averages and displacement of deaths by a few weeks will not influence these), these results can imply important public health effects of the changing climate. In addition to the seasonal mean temperature, its variability may also play a significant role in raising the risk of mortality\textsuperscript{12,14}. There is evidence that people adapt to the usual temperature in their city\textsuperscript{14,15}. For these reasons, this study focused on both mean temperature, and temperature variability for summer and winter.

Conventionally, ambient temperature ($T_a$) was obtained from monitors near airports, and analyzed at the city or county level\textsuperscript{16}. Overlooking the temporal and spatial variation of the $T_a$ may introduces an exposure measurement error and may bias health effect estimates\textsuperscript{17}. We recently presented new hybrid models of assessing high-resolution spatio-temporal $T_a$ for epidemiological studies, based on surface temperature ($T_s$) measured by satellite. This approach allows prediction of daily $T_a$ at $1 \times 1$ km spatial resolution throughout the New England area\textsuperscript{18}. We also showed that using high-resolution $T_a$ predicted from the hybrid models better captured the associations between $T_a$ and adverse health outcomes than temperature data from existing monitoring stations\textsuperscript{19}.

Using temporally- and spatially-resolved $T_a$ estimates, the present study aims to tease apart the associations of annual all-cause mortality with seasonal mean temperature and with temperature variability for both summer and winter, among the Medicare population in New England during 2000-2008, and to further separate these into long
term spatial differences and annual anomalies. Temperature variability is represented by
the standard deviation (SD) of daily mean temperature within season.

The results, presented as percent increase in mortality, are shown in Fig. 1 for
spatial contrasts, annual anomalies, and the overall effect estimates. Both seasonal mean
temperatures and temperature SDs were significantly associated with all-cause mortality
for summer and winter ($p<0.05$). A rise of 1 °C in summer mean temperature
corresponded to an overall estimated 1.0% increase in mortality (95% CI: 0.6, 1.5%).
This was entirely due to yearly anomalies; living in a location with long-term warmer
summer temperatures in New England was associated with lower mortality rates,
suggesting acclimatization. A 1 °C increase in winter mean temperature was observed to
lower mortality by 0.6% (95% CI: 0.3, 0.9%). Here, it was the spatial contrast that was
associated with lower mortality, and the anomalies with higher mortality, suggesting little
acclimatization to cold. For each 1 °C increase in temperature SDs, 1.3% (95% CI: 0.2,
2.4%) and 4.1% (95% CI: 3.0, 5.2%) increases in annual deaths were seen in summer and
winter, respectively. The observed associations in summer and winter were respectively
attributable to spatial contrasts and yearly anomalies.
Figure 1: Percent increase in mortality for per 1 °C and per Interquartile Range (IQR) increases in seasonal mean temperature and temperature variability across New England, 2000-2008. IQR is defined as; IQR = third quartile – first quartile which reflects the current distribution of these variables. A sensitivity analysis controlling for the number of heat waves in each ZIP code in each year found essentially identical results. Error bars stand for 95% confidence intervals. * p < 0.05.

A key result of this study is that increases in temperature variability can have similar estimated effects on mortality as increases in mean temperature. This makes understanding the effect of climate change on seasonal temperature variability important.
for assessing health risks. In addition, we found that small area variations in long term average temperature and temperature SD across New England are associated with differences in mortality rates, which are quite different from the associations with anomalies. In particular they suggest even within New England, places with typically warmer summers have lower annual mortality rates while anomalies are associated with higher annual mortality rates in that year. Hence the negative health impacts of climate change may stem from the increase in year-to-year fluctuations of temperature from a level that people have been acclimated to for a long time, or an increase in daily variability within season, rather than the warmer temperature itself.

Short-term associations between daily temperature and mortality have been reported to be greater in locations with larger seasonal temperature variability9,14. Several recent studies suggested that a large change in temperature between neighboring days can elevate daily mortality10,11. Hence temperature variability emerges as an important risk factor for both short-term and long-term survival.

Subjects exercising in hot conditions for 12 days doubled their endurance time, while subjects alternately exercising in mild and hot conditions did not20. This suggests that more temperature variability in the summer can impede adaptation, which is consistent with our associations with seasonal SD. However, the underlying mechanisms for the observed relationship with seasonal mean temperature remain unresolved. That more people die when summers are warmer or when winters are colder than normal for their ZIP code may be due to accumulating effects over the season on the changes in blood pressure associated with short-term change in temperature, or lack of behavioral response.
Our results also suggest that living in warmer ZIP codes in New England reduced long-term mortality in both summer and winter, but living in ZIP codes with more variable temperature in either summer or winter increased mortality. In addition, higher annual anomalies were significantly associated with increased long-term mortality, though insignificantly for summer temperature SD.

We also examined effect modifications (Fig. 2 and 3). For seasonal mean temperature anomalies, consistent increases in mortality rates for positive anomalies were seen across the groups (cf. Fig. 2). While some differences were statistically significant (e.g., between males and females for summer mean), the differences in effects were not large except for some noise in smaller racial groups. Anomalies for the SD of temperature within winter also showed consistently positive associations across group. Summer temperature SD anomalies showed no association with mortality overall, but opposite effects by age, with higher mortality rates in persons 75+, and lower rates in 65-75.
Figure 2: Modifications of the effect for annual anomalies of different temperature indices by population density, age, sex, and race. It displays the percent increases of death (95% CI) for each 1 °C increase in (a) summer mean temperature annual anomaly (b) winter mean temperature annual anomaly (c) summer temperature standard deviation annual anomaly (d) winter temperature standard deviation annual anomaly in each subgroup, respectively. Red indicates increased risk of death, and blue stands for decreased risk of death. Error bars stand for 95% confidence intervals. * $p < 0.05$. 
For ZIP code differences in the long term means of the temperature parameters, more heterogeneity was seen across population groups (cf. Fig. 3). Higher effect estimates were found in less dense areas for both mean and SD of summer temperature. This is consistent with a recent study suggesting that subjects in rural counties were more vulnerable to heat\textsuperscript{21}. We also found that subjects in urban areas benefited more from living in regions with warmer winters but were at higher risk for elevated winter temperature variability. In addition, the relatively younger age group (65-74 y) was more sensitive to the changes of most temperature indices. This may be due to their having more outdoor activities than the older age group. A similar reason may also explain why in summer higher effect estimates were found for subjects living in rural areas. However in the cold and snowy winter of New England, possibly outdoor activities of studied subjects in rural areas were limited, therefore winter mean temperature and SD had lower effect estimates for subjects in rural areas than that in urban. Females were more sensitive to increased mean temperature in summer and winter but less sensitive to increased summer temperature SD compared to males. We saw a number of interactions by race that are not always easy to understand. It is possible that the decreased risk of death in locations with higher summer mean temperatures in Asian, Hispanic and “other” race group is because elderly subjects in these groups immigrated from warm areas, such as South China, Southeast Asia, India, and Mexico. These subjects were more acclimated to a warmer summer. This could also explain that Asian, Hispanic and “other” race groups significantly benefited more from a warmer winter, but were more adversely impacted by greater winter SD. However, we note that the percent of the population over 65 in New
England that is not white is only 5.9%, and for 4.3% ZIP codes was zero, so some of the observed variability across race may result from unstable estimates.

Figure 3: Modifications of the effect for spatial variability in different temperature indices by population density, age, sex, and race. It displays the percent increases of death (95% CI) for each 1 °C increase in (a) spatial contrasts of summer mean temperature (b) spatial contrasts of winter mean temperature (c) spatial contrasts of summer temperature standard deviation (d) spatial contrasts of winter temperature standard deviation in each subgroup, respectively. Red indicates increased risk of death,
and blue stands for decreased risk of death. Error bars stand for 95% confidence intervals. * $p < 0.05$.

A potential limitation of the present study is that there is still potential residual confounding resulting from the lack of individual level risk factors for mortality. In addition, bias could also be induced by spatial or temporal auto-correlation, which can lead to spuriously narrow standard errors and CIs.

**Implications for future climate research**

Scientific understanding of how mean surface temperature responds to the increasing anthropogenic emission of GHGs has been greatly improved in recent decades\(^1\). Considering the comparable increments of both summer and winter mean temperatures predicted for the future 50 years, the benefit of warmer winter may be largely compensated by the harm of hotter summer. However, for annual anomalies, warmer winter temperature actually increases annual mortality. The degree of temperature variability, such as the within season SD, which is influenced by many factors such as the intensity and pattern of atmospheric circulations, remains highly uncertain for future climate scenarios\(^22,23\). For example, Petoukhov and Semenov proposed that the weakening of the polar vortex may bring about increased frequency of very cold winters in northern America\(^24\). However, Huntingford and coauthors suggested that there is no significant change in global temperature variability in spite of changing regional patterns\(^3\). That study assessed variability across the entire year, including
seasonal variability, and it is not clear whether variability changed within some seasons \(^3\). This uncertainty can be propagated into the projection of temperature-related health risk.

This study, by estimating impacts of both mean temperature and temperature variability in summer and winter simultaneously, adds considerable strength to the evidence of a significant association between prolonged exposures to temperatures and mortality, especially temperature variability. In addition to the within season variability we also showed that between year variability (anomalies) increased mortality risks. Hence the variability of atmospheric temperature emerges as a key factor of the potential health impacts of climate change.

This work provides an important example of how temperature may affect human health in a temperate climate region. We would expect that the health effects of seasonal mean temperature and temperature variability can vary greatly among different climate zones \(^1\). Such variation may also exist across areas with distinct socio-economical status, which can convey vulnerability to the changing climate. A quantitative assessment of the projected risk of human health associated with future climate change can be estimated world-wide by incorporating more comprehensive epidemiological studies and projected climate scenario data for different climate zones. However, as an important environmental stressor emphasized in this study, the uncertainty of the projected changes in the temperature variability can induce significant errors in such estimate. Hence better climate projections on temperature variability at multiple scales are important for the human health risk assessment.
Methods

Study Population. We constructed a cohort using Medicare data for all residents aged 65 and older for the years 2000-2008 in New England (cf. Table S1 in Supplemental Material). This was an open cohort, including eligible persons from 1999, or at the year when they subsequently turned 65. Subjects entered the cohort for survival if they were still alive on January 1st of the year following the year that they enrolled in Medicare, and follow-up years (our time metric) were calendar years. As a national social insurance program administered by the U.S. federal government since 1966, Medicare guarantees Americans aged 65 and older access to health insurance. The Medicare beneficiary denominator file from the Centers for Medicare and Medicaid services (CMS) lists all beneficiaries enrolled in the Medicare fee-for-service (FFS), and contains information on beneficiaries’ eligibility and enrollment in Medicare, as well as the date of death.

Exposure data. The present study uses 1 × 1 km ambient temperature (T_a) data estimated from surface temperature (T_s) measured by satellites. Specifically, we started by calibrating the T_s - T_a relationship for each day using grid cells with both T_a and T_s measurements (model 1). This daily calibration was then used to predict T_a in grid cells in the study domain without T_a measurements but with available T_s measurements. To fill in cells or days when no T_s measurements were available, a generalized additive mixed model was fit with a smooth function of latitude and longitude of the grid cell centroid (model 2). The performance of the estimated T_a was validated by ten-fold cross-validation. Out-of-sample R^2 was found to be very high (R^2 = 0.947, yearly variation 0.933-0.958 for the years 2000-2011) for days with available satellite T_s measurements. Excellent performance was also observed even in days with no available T_s data (R^2 =
0.940, yearly variation 0.902-0.962 for the years 2000-2011). More details are published\textsuperscript{17}.

By linking the ZIP code centroid to the nearest temperature grid, we assigned the grid-cell temperature exposures to each ZIP code. The predicted daily ambient temperatures allow us to calculate, for each year and each ZIP code, the mean temperature for summer (June-August), the mean temperature for winter (December-February), the standard deviation (SD) of daily mean summertime temperature, and the SD of daily mean wintertime temperature. We refer to these as summer mean temperature, winter mean temperature; summer temperature variability and winter temperature variability, respectively (\textit{cf. Table S2 in Supplemental Material}). Table S2 also presents the distribution of the spatial and temporal variations for the temperature variables. Table S3 in Supplemental Material displays the correlation coefficients between these explanatory variables.

\textbf{Covariates.} Medicare provides information on age, race and sex of all individuals. From the U.S. Census Bureau 2000, we obtained ZIP code tabulation area level socio-economic variables, including population density, percent of green space, percent of the population (age\textgreater=65) in poverty status, and median value for owner occupied housing units. In addition, based on ZIP code level primary and secondary hospital admissions for lung cancer, we calculated the long-term average hospital admission rate for lung cancer in each ZIP code as a surrogate for smoking experience. The census data were merged with individuals based on their ZIP code of residence. County-level percent of diabetes and percent of lack of physical activity, obtained from the CDC Behavioral Risk Factor Surveillance survey for the entire country were adjusted as well.
Statistical Methods. In our dataset, one observation is created for each Medicare participant for each year of follow-up, using the Andersen Gill formulation of survival analysis. Survival times were calculated from enrollment date until death or December 31, 2008 (censoring), whichever came first.

We considered the following possible exposure indices: summer mean temperature, winter mean temperature, summertime temperature SDs and wintertime temperature SDs in each follow-up year. To separate the independent associations of mortality with mean temperature and temperature variability, all temperature-related indices were entered into the models simultaneously and treated as time-varying exposures.

We applied extended Cox proportional hazard models (Proc PHREG, SAS 9.3), which allow for time-varying covariates in survival analysis. The models were adjusted for individual risk factors including age, sex, race, ZIP code level covariates such as population density, percent of green space, percent of the population below poverty level, median value for owner occupied housing units, and hospital admission rate for lung cancer (a surrogate for smoking experience), as well as county-level percent of diabetes, and percent of lack of physical activity. To adjust for time trend, we entered an indicator for each year of follow-up. To allow for possible non-proportionality of hazard, participants were stratified by sex, 5-year age groups, and race (white, black, and others), such that each sex/age/race group had its own baseline hazard.

We then assessed the association of mortality with the four exposure indices. The analyses were also repeated without mutual adjustment for seasonal mean temperatures and temperature SDs (cf. Table S4 in Supplemental Material). To separate the
contribution of the spatial and temporal components of the temperature variables, we fit a model with two terms for each temperature variable. We defined each term as its mean in that ZIP code over the entire study period plus that difference (anomaly) from that mean for \( i \)th ZIP code in \( t \)th follow-up year. With all 8 terms put in the model, the mean for each ZIP code captures purely geographic contrasts, the anomaly indicate the effects of yearly variations within each ZIP code.

The results are expressed as percent increase in mortality for each °C increase in an exposure index, for spatial contrasts, annual anomalies, and the overall effect estimates. Percent increases in mortality were also calculated for per IQR (Interquartile Range; IQR = third quartile – first quartile) increase, to reflect the current distribution of these variables. For example, an 8.4% (95% CI: 7.3, 9.6%) increase of mortality was observed for per IQR increase of yearly anomaly of summer mean temperature. This result can be interpreted as that within a ZIP code, the risk of death in a relatively warm summer (third quartile in the studied period) was 8.4% higher than in a relatively cold summer (first quartile). A sensitivity analysis controlling for the number of heat waves in each ZIP code in each year was performed as well.

We also examined age (defined as >=75 y and 65-74 y), sex, race (white, black, Asian, Hispanic, and other), and population density (less than 33th percentile, rural, and otherwise urban) as effect modifiers for both spatial contrasts and temperature anomalies respectively, by adding interaction terms between such variables and each of the eight temperature indices in the model.
Table S1. Characteristics of the study population across New England for the years 2000-2008

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total population</strong></td>
<td>2740308</td>
<td>100</td>
</tr>
<tr>
<td><strong>Deaths</strong></td>
<td>836853</td>
<td>30.5</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1158052</td>
<td>42.3</td>
</tr>
<tr>
<td>Female</td>
<td>1582256</td>
<td>57.7</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>2576263</td>
<td>94.0</td>
</tr>
<tr>
<td>Black</td>
<td>83444</td>
<td>3.1</td>
</tr>
<tr>
<td>Asian</td>
<td>22005</td>
<td>0.8</td>
</tr>
<tr>
<td>Hispanic</td>
<td>19575</td>
<td>0.7</td>
</tr>
<tr>
<td>Other</td>
<td>39021</td>
<td>1.4</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65-69</td>
<td>611977</td>
<td>22.3</td>
</tr>
<tr>
<td>70-74</td>
<td>521568</td>
<td>19</td>
</tr>
<tr>
<td>75-79</td>
<td>514346</td>
<td>18.8</td>
</tr>
<tr>
<td>80-84</td>
<td>476469</td>
<td>17.4</td>
</tr>
<tr>
<td>85-89</td>
<td>353639</td>
<td>12.9</td>
</tr>
<tr>
<td>&gt;=90</td>
<td>262309</td>
<td>9.6</td>
</tr>
</tbody>
</table>
Table S2. Descriptive Statistics for temperature and temperature variability in summer and winter across New England, 2000-2008

<table>
<thead>
<tr>
<th>Temperature (°C)</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Range</th>
<th>IQR</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer mean</td>
<td>16.39</td>
<td>16.34</td>
<td>1.16</td>
<td>1.75</td>
<td>33.34</td>
<td>31.59</td>
<td>1.60</td>
<td>15.57</td>
<td>17.17</td>
</tr>
<tr>
<td>Winter mean</td>
<td>2.35</td>
<td>2.68</td>
<td>2.88</td>
<td>-7.47</td>
<td>12.20</td>
<td>19.67</td>
<td>4.03</td>
<td>0.38</td>
<td>4.41</td>
</tr>
<tr>
<td>Summer SD</td>
<td>2.46</td>
<td>2.45</td>
<td>0.33</td>
<td>0.00</td>
<td>4.50</td>
<td>4.50</td>
<td>0.44</td>
<td>2.23</td>
<td>2.67</td>
</tr>
<tr>
<td>Winter SD</td>
<td>3.55</td>
<td>3.48</td>
<td>0.64</td>
<td>0.00</td>
<td>7.56</td>
<td>7.56</td>
<td>0.85</td>
<td>3.10</td>
<td>3.95</td>
</tr>
<tr>
<td>Spatial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer mean</td>
<td>16.49</td>
<td>16.48</td>
<td>0.99</td>
<td>13.37</td>
<td>19.19</td>
<td>5.82</td>
<td>1.32</td>
<td>15.80</td>
<td>17.11</td>
</tr>
<tr>
<td>Winter mean</td>
<td>2.58</td>
<td>2.95</td>
<td>2.24</td>
<td>-3.98</td>
<td>8.22</td>
<td>12.19</td>
<td>3.40</td>
<td>0.96</td>
<td>4.36</td>
</tr>
<tr>
<td>Summer SD</td>
<td>2.46</td>
<td>2.46</td>
<td>0.14</td>
<td>0.85</td>
<td>3.01</td>
<td>2.16</td>
<td>0.14</td>
<td>2.39</td>
<td>2.53</td>
</tr>
<tr>
<td>Winter SD</td>
<td>3.41</td>
<td>3.29</td>
<td>0.46</td>
<td>1.50</td>
<td>5.06</td>
<td>3.56</td>
<td>0.63</td>
<td>3.09</td>
<td>3.72</td>
</tr>
<tr>
<td>Temporal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer mean</td>
<td>0.01</td>
<td>0.00</td>
<td>0.59</td>
<td>-6.21</td>
<td>15.61</td>
<td>21.83</td>
<td>0.88</td>
<td>-0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>Winter mean</td>
<td>-0.01</td>
<td>-0.06</td>
<td>1.45</td>
<td>-5.90</td>
<td>4.96</td>
<td>10.85</td>
<td>2.03</td>
<td>-1.03</td>
<td>1.00</td>
</tr>
<tr>
<td>Summer SD</td>
<td>0.00</td>
<td>0.01</td>
<td>0.26</td>
<td>-0.85</td>
<td>2.86</td>
<td>3.71</td>
<td>0.37</td>
<td>-0.19</td>
<td>0.18</td>
</tr>
<tr>
<td>Winter SD</td>
<td>0.00</td>
<td>0.07</td>
<td>0.50</td>
<td>-1.71</td>
<td>3.91</td>
<td>5.62</td>
<td>0.67</td>
<td>-0.31</td>
<td>0.36</td>
</tr>
</tbody>
</table>
Table S3. Correlation coefficients ($r$) of the spatial and temporal variations for mean temperature and temperature variability

<table>
<thead>
<tr>
<th></th>
<th>Summer mean</th>
<th>Winter mean</th>
<th>Summer SD</th>
<th>Winter SD</th>
<th>Summer mean$'$</th>
<th>Winter mean$'$</th>
<th>Summer SD$'$</th>
<th>Winter SD$'$</th>
<th>Summer mean$''$</th>
<th>Winter mean$''$</th>
<th>Summer SD$''$</th>
<th>Winter SD$''$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer mean</td>
<td>1.00</td>
<td>0.56</td>
<td>0.43</td>
<td>-0.37</td>
<td>0.86</td>
<td>0.69</td>
<td>0.56</td>
<td>-0.56</td>
<td>0.51</td>
<td>-0.04</td>
<td>0.19</td>
<td>0.01</td>
</tr>
<tr>
<td>Winter mean</td>
<td>0.56</td>
<td>1.00</td>
<td>0.12</td>
<td>-0.82</td>
<td>0.68</td>
<td>0.84</td>
<td>0.18</td>
<td>-0.80</td>
<td>-0.05</td>
<td>0.54</td>
<td>0.03</td>
<td>-0.37</td>
</tr>
<tr>
<td>Summer SD</td>
<td>0.43</td>
<td>0.12</td>
<td>1.00</td>
<td>-0.02</td>
<td>0.31</td>
<td>0.10</td>
<td>0.48</td>
<td>-0.02</td>
<td>0.32</td>
<td>0.06</td>
<td>0.88</td>
<td>-0.01</td>
</tr>
<tr>
<td>Winter SD</td>
<td>-0.37</td>
<td>-0.82</td>
<td>-0.02</td>
<td>1.00</td>
<td>-0.44</td>
<td>-0.64</td>
<td>-0.03</td>
<td>0.67</td>
<td>0.01</td>
<td>-0.51</td>
<td>-0.00</td>
<td>0.74</td>
</tr>
<tr>
<td>Summer mean$'$</td>
<td>0.86</td>
<td>0.68</td>
<td>0.31</td>
<td>-0.44</td>
<td>1.00</td>
<td>0.80</td>
<td>0.65</td>
<td>-0.65</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Winter mean$'$</td>
<td>0.69</td>
<td>0.84</td>
<td>0.10</td>
<td>-0.64</td>
<td>0.80</td>
<td>1.00</td>
<td>0.22</td>
<td>-0.95</td>
<td>-0.01</td>
<td>-0.00</td>
<td>-0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Summer SD$'$</td>
<td>0.56</td>
<td>0.18</td>
<td>0.48</td>
<td>-0.03</td>
<td>0.65</td>
<td>0.22</td>
<td>1.00</td>
<td>-0.05</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Winter SD$'$</td>
<td>-0.56</td>
<td>-0.80</td>
<td>-0.02</td>
<td>0.67</td>
<td>-0.65</td>
<td>-0.95</td>
<td>-0.05</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.01</td>
</tr>
<tr>
<td>Summer mean$''$</td>
<td>0.51</td>
<td>-0.05</td>
<td>0.32</td>
<td>0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>-0.08</td>
<td>0.36</td>
<td>0.02</td>
</tr>
<tr>
<td>Winter mean$''$</td>
<td>-0.04</td>
<td>0.54</td>
<td>0.06</td>
<td>-0.51</td>
<td>0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.08</td>
<td>1.00</td>
<td>0.07</td>
<td>-0.69</td>
</tr>
<tr>
<td>Summer SD$''$</td>
<td>0.19</td>
<td>0.03</td>
<td>0.88</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.36</td>
<td>0.07</td>
<td>1.00</td>
<td>-0.01</td>
</tr>
<tr>
<td>Winter SD$''$</td>
<td>0.01</td>
<td>-0.37</td>
<td>-0.01</td>
<td>0.74</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.02</td>
<td>-0.69</td>
<td>-0.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>

$^s$ stands for spatial variation;
$t$ stands for temporal variation.
### Table S4. Percent increase in mortality (95% CI) for each 1 °C increase for both seasonal mean temperature and temperature SD

<table>
<thead>
<tr>
<th>Variable</th>
<th>Percent increase</th>
<th>95% CI</th>
<th>p-value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unadjusted by SDs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer mean</td>
<td>1.7</td>
<td>(1.3,2.1)</td>
<td>&lt;.0001</td>
<td>1 °C</td>
</tr>
<tr>
<td>Winter mean</td>
<td>-1.6</td>
<td>(-1.7,-1.4)</td>
<td>&lt;.0001</td>
<td>1 °C</td>
</tr>
<tr>
<td><strong>Unadjusted by means</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer SD</td>
<td>2.2</td>
<td>(1.2,3.2)</td>
<td>&lt;.0001</td>
<td>1 °C</td>
</tr>
<tr>
<td>Winter SD</td>
<td>5.6</td>
<td>(5.0,6.2)</td>
<td>&lt;.0001</td>
<td>1 °C</td>
</tr>
<tr>
<td><strong>Adjusted by SDs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer mean</td>
<td>1.0</td>
<td>(0.6,1.5)</td>
<td>&lt;.0001</td>
<td>1 °C</td>
</tr>
<tr>
<td>Winter mean</td>
<td>-0.6</td>
<td>(-0.9,-0.3)</td>
<td>&lt;.0001</td>
<td>1 °C</td>
</tr>
<tr>
<td><strong>Adjusted by means</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer SD</td>
<td>1.3</td>
<td>(0.2,2.4)</td>
<td>&lt;.0001</td>
<td>1 °C</td>
</tr>
<tr>
<td>Winter SD</td>
<td>4.1</td>
<td>(3.0,5.2)</td>
<td>&lt;.0001</td>
<td>1 °C</td>
</tr>
</tbody>
</table>
REFERENCES

1 IPCC. Climate change 2013: The physical science basis. (Cambridge University Press Cambridge, UK, and New York, 2013).


CHAPTER III

Low-Concentration PM\textsubscript{2.5} and Mortality: Estimating Acute and Chronic Effects in a Population Based Study

Liuhua Shi\textsuperscript{1}, Antonella Zanobetti\textsuperscript{1}, Itai Kloog\textsuperscript{1,2}, Brent A. Coull\textsuperscript{3}, Petros Koutrakis\textsuperscript{1}, Steven J. Melly\textsuperscript{1}, Joel D. Schwartz\textsuperscript{1}

\textsuperscript{1}Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA

\textsuperscript{2}Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Beer Sheva, Israel

\textsuperscript{3}Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
ABSTRACT

Background: Both short- and long-term exposures to fine particulate matter (PM$_{2.5}$) are associated with mortality. However, whether the associations exist below the new EPA standards (12 μg/m$^3$ of annual average PM$_{2.5}$, 35 μg/m$^3$ daily) is unclear. In addition, it is not clear whether results of previous time series studies (fit in larger cities) and cohort studies (fit in convenience samples) are generalizable to the general population.

Objectives: To estimate the effects of low-concentration PM$_{2.5}$ on mortality.

Methods: High resolution (1 × 1 km) daily PM$_{2.5}$ predictions, derived from satellite aerosol optical depth retrievals, were employed. Poisson regressions were applied to the Medicare population (age>=65) in New England to simultaneously estimate the acute and chronic effects, with mutual adjustment for short- and long-term exposure, as well as area-based confounders. Models were also restricted to annual concentrations below 10 μg/m$^3$ or daily concentrations below 30 μg/m$^3$.

Results: PM$_{2.5}$ was associated with increased mortality. In the cohort, 2.14% (95% CI: 1.38, 2.89%) and 7.52% (95% CI: 1.95, 13.40%) increases were estimated for each 10 μg/m$^3$ increase in short- (2 day) and long-term (1 year) exposures, respectively. The associations still held for analyses restricted to low-concentration PM$_{2.5}$ exposures. The corresponding estimates were 2.14% (95% CI: 1.34, 2.95 %) and 9.28% (95% CI: 0.76, 18.52%). Penalized spline models of long-term exposure indicated a higher slope for mortality in association with exposures above versus below 6 μg/m$^3$. In contrast, the association between short-term exposure and mortality appeared to be linear across the entire exposure distribution.
**Conclusions:** Using a mutually-adjusted model, we estimated significant acute and chronic effects of PM$_{2.5}$ exposures below current EPA standards. These findings suggest that improving air quality below current standards may benefit public health.
INTRODUCTION

Many studies have found association between fine particulate matter (PM with aerodynamic diameter \( \leq 2.5 \mu m \) \([\text{PM}_{2.5}]\)) and increased mortality (Dockery et al. 1993; Schwartz 1994; Pope III et al. 2002; Franklin et al. 2007; Zanobetti and Schwartz 2009). Biological evidence has been established for plausible mechanisms between \([\text{PM}_{2.5}]\) and mortality, such as increased risk of ventricular arrhythmia and thrombotic processes, increased system inflammation and oxidative stress, increased blood pressure, decreased plaque stability, and lower lung function, among others (Downs et al. 2007; Gauderman et al. 2004; Suwa et al. 2002; Brook et al. 2009; Gurgueira et al. 2002; Yue et al. 2007).

Based on evidence from epidemiological and toxicological studies (Chen and Nadziejko 2005; Furuyama et al. 2006; Ohtoshi et al. 1998), National Ambient Air Quality Standards (NAAQS) were implemented for fine particulate matter. For example, U.S. Environmental Protection Agency (EPA) revised the fine particle NAAQS in the years of 1997, 2006, and 2012, in order to protect public health (EPA 1997; EPA 2006; EPA 2013). Further changes in the standards require additional studies to elucidate whether health effects occur at levels below current annual and daily EPA NAAQS of 12 and 35 \( \mu g/m^3 \), respectively. The Clean Air Act requires EPA to review national air quality standards every five years to determine whether they should be retained or revised (United States Code Title 42, Chapter 85), thus whether the relationship continues below the current standards is of great interest and importance.

Previous studies have generally focused on either long-term (Hart et al. 2011; Puett et al. 2009; Schwartz 2000; Jerrett et al. 2005) or short-term (Schwartz and Dockery 1992; Katsouyanni et al. 1997; Dominici et al. 2006; Samoli et al. 2008) exposures across
the entire range of PM$_{2.5}$ concentration. In the case of time series analyses of short-term exposures, the need to ensure the relevance of the monitoring data and sufficient population for power has limited analyses to larger cities, and hence exurbs, smaller cities, and rural areas are not generally represented in the literature. This may compromise the generalizability of the results. In addition, there is spatial variability in PM$_{2.5}$ concentrations within cities that time series studies generally have not taken into account, introducing exposure measurement error (Lepeule et al. 2012; Laden et al. 2006).

Chronic effects studies began using comparisons across cities of mortality experiences of a cohort living in various communities and the monitored air pollutant concentrations in those communities (Dockery et al. 1993; Pope et al. 1995). Again, these suffered from exposure error due to failure to capture within city spatial variability in exposure. Since the geographic exposure gradient is the exposure contrast in these studies, the failure to capture within city contrasts leads to classical measurement error, with expected downward bias. Studies with e.g., land use regression estimates of exposure have generally reported larger effect sizes (Puett et al. 2009; Miller et al. 2007). Previous cohort studies have not controlled for the acute effects of particles when estimating chronic effects, raising the question of whether there are independent chronic effects that represent more than the cumulative effects of acute responses.

In general, existing study cohorts are not representative of the population. For example, the American Cancer Society (ACS) cohort has higher education than the US population as a whole (Stellman and Garfinkel 1986). Hence, there have been few population-based cohort studies conducted until recently (Kloog et al. 2013).
Several time series studies examined the concentration-response relationship between PM$_{2.5}$ and mortality below concentrations of 100 µg/m$^3$. They generally report linear concentration-response (Samoli et al. 2008; Schwartz and Zanobetti 2000). However, there have been few studies focusing on exposures below the current daily EPA standard of 35 µg/m$^3$.

More studies have examined the shape of the concentration-response curve for long-term exposure versus short-term exposure, but they have mostly not covered population-based cohorts, or included very low exposures (Schwartz et al. 2008; Crouse et al. 2012).

We recently presented a new hybrid method of assessing temporally- and spatially-resolved PM$_{2.5}$ exposure for epidemiological studies by incorporating 1 × 1 km resolution satellite-retrieved Aerosol optical Depth (AOD) measurements, combined with traditional land use terms, meteorological variables and their interactions (Kloog et al. 2014a). This approach allows for predicting daily PM$_{2.5}$ concentrations at a 1 × 1 km spatial resolution throughout the New England area. We also validated our models performance in rural areas: Ten-fold cross-validation of our model in rural areas (using the IMPROVE stations) resulted in a CV $R^2$ of 0.92. Further details have been published (Kloog et al. 2014a).

The present study aims to simultaneously estimate acute and chronic health effects of PM$_{2.5}$ in a population-based Medicare cohort (age ≥ 65) encompassing the New England region. It uses high spatial resolution exposure estimates based on satellite measurements that are available across the region and not just in limited locations. To make this study relevant to future assessments of current EPA standards, we repeated the analysis after restricting the data to long-term exposures (365-day moving average) below 10 µg/m$^3$. 

---

58
and repeated the time series analysis of short-term exposures after restricting data to two day average exposures below 30 $\mu$g/m$^3$.

**METHODS**

**Study domain**

The spatial domain of our study included the New England area, comprising the states of Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island and Vermont (Figure 1A).

![Figure 1A](image1.png)

**Figure 1A.** Mean PM$_{2.5}$ concentrations in 2004 at a high resolution ($1 \times 1$ km) across New England predicted by the AOD models.
Exposure data

Novel models for predicting daily PM$_{2.5}$ were previously reported incorporating AOD and land use data for the New England region (Kloog et al. 2011). Previous papers have shown that using actual physical measurements in our prediction models improves predictive accuracy over comparable land use or spatial smoothing models (Kloog et al. 2011). With AOD retrieved by Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm, a similar approach was applied for estimating daily PM$_{2.5}$ exposures in New England at a spatial resolution of 1 × 1 km (Kloog et al. 2014a). In this study, the same PM$_{2.5}$ exposure predictions were employed.

In brief, we calibrated the AOD-PM$_{2.5}$ relationship on each day during the study period (2003-2008) using data from grid cells with both ground PM$_{2.5}$ monitors and AOD measurements (stage 1), and used inverse probability weighting to address selection bias due to non-random missingness patterns in the AOD measurements. We then used the AOD-PM$_{2.5}$ relationship to predict PM$_{2.5}$ concentrations for grid cells that lacked monitors but had available AOD measurement data (stage 2). Finally, we used a generalized additive mixed model (GAMM) with spatial smoothing and a random intercept for each 1 × 1 km grid cell to impute data for grid cell/days when AOD measurements were not available (stage 3). The performance of the estimated PM$_{2.5}$ was validated by ten-fold cross-validation. High out-of-sample R$^2$ (R$^2$=0.89, year to year variation 0.88 - 0.90 for the years 2003-2008) was found for days with available AOD data. Excellent performance held even in cells/days with no available AOD (R$^2$=0.89, year to year variation 0.87 - 0.91 for the years 2003-2008). The 1 km model has better
spatial (0.87) and temporal (0.87) out of sample $R^2$ than the previous 10 km model (0.78 and 0.84 respectively). Details of the PM$_{2.5}$ prediction models are found in Kloog et al (Kloog et al. 2014a).

Figure 1A shows an example of mean PM$_{2.5}$ concentrations in 2004 at a $1 \times 1$ km spatial resolution across New England. By averaging our estimated daily exposures at each location we generated long-term exposures.

Figure 1B. Predicted PM$_{2.5}$ concentrations at a $1 \times 1$ km grid for November 15, 2003.

Figure 1B (a subset of the study area) shows that spatial variability exists even for daily data, and is not identical to the long-term pattern shown in Figure 1A. That is, there is space-time variation in the PM$_{2.5}$ exposure captured in this analysis, but not in previous time series analyses.
Because the deaths were coded at the ZIP code level, both long- and short-term predictions were matched to ZIP codes by linking the ZIP code centroid to nearest PM$_{2.5}$ grid, using ArcGIS (ESRI, Redlines, CA) and SAS (SAS Institute, Cary, NC).

Traditionally, studies of the acute air pollution effects have controlled for temperature using values taken from the nearest airport. This approach is not feasible for the entire region, since too many residences are distant from airports. In addition, there is spatio-temporal variation in temperature as well. We have used similar models of satellite data to fit models for temperature in New England on a 1 km temperature data estimated from surface temperature measured by satellites (Kloog et al. 2014b). To our knowledge, such fine control for temperature has not previously been used in air pollution epidemiology.

**Mortality data**

Individual mortality records were obtained from the US Medicare program for all residents aged 65 and older for all available years during 2003-2008 (CMS). The MEDICARE cohort was used because of availability of ZIP code of residence data, whereas National Center for Health Statistics mortality data is only available on a county level. Additionally, prior studies found that elderly people are more susceptible to the effects of particulate matter (Pope 3rd 2000). The Medicare beneficiary denominator file from the Centers for Medicare and Medicaid services (CMS) lists all beneficiaries enrolled in the Medicare fee-for-service (FFS), and contains information on beneficiaries’ eligibility and enrollment in Medicare, and the date of death (CMS). The Medicare Provider Analysis and Review (MEDPAR) inpatient data includes information on age, gender, race, ZIP code of residence, and one record for each hospital admission (CMS).
Daily mortality was first aggregated by ZIP code, and then matched with the corresponding PM$_{2.5}$ exposure. We summarized the mortality data by ZIP code and day because that is the finest resolution of address we can obtain. Since the mortality datasets did not include changes of residence, we had to assume that the subjects lived at their current address over the study period.

**Covariates**

We used daily 1 km temperature data estimated from surface temperature measured by satellites (Kloog et al. 2014b). All socio-economic variables were obtained through the U.S. Census Bureau 2000 Census Summary File 3, which includes social, economic, and housing characteristics [United States Census Bureau (USCB) 2000]. ZIP code tabulation area level socio-economic variables, including race, education, and median household income, were used. In addition, county-level percent of people who currently smoke every day from the CDC Behavioral Risk Factor Surveillance survey for the entire country was adjusted as well (CDC 2013). Dummy variables were used to control for day of the week.

**Statistical analysis**

Conventionally, the acute effects of air pollution are estimated by Poisson log-linear models and the chronic effects are estimated by Cox proportional hazard models (Laden et al. 2006; Kloog et al. 2013). Laird and Olivier pointed out the equivalence of the likelihood of a proportional hazard model with piecewise constant hazard for each year of follow-up and a Poisson regression with a dummy variable for each year of follow-up (Laird and Olivier 1981). We have taken advantage of this to generalize from dummy variables for each year to a spline of time to represent the baseline hazard, and to
aggregate subjects into counts per person time at risk and obtained a mixed Poisson regression model (Kloog et al. 2012). This approach allows one to model the rate of death as a function of both long- and short-term exposures simultaneously. By doing so, we achieve the equivalence of a separate time series analysis for each ZIP code, greatly reducing the exposure error in that part of the model, while simultaneously conducting a survival analysis on the participants, and also are able to estimate the independent effects of both exposures.

Most time series studies have reported stronger associations with acute exposures when exposures are defined as mean PM$_{2.5}$ on the day of death and the previous day (lag01), compared with mean PM$_{2.5}$ on the current day only, or for exposures with longer lags (Schwartz et al. 1996; Schwartz 2004). We used the lag01 average for our main analysis, but performed a sensitivity analysis on that choice. Long-term exposure was calculated as the 365-day moving average ending on date of death so that our results were comparable with previous studies (Lepeule et al. 2012; Schwartz et al. 2008). Short-term exposure was defined as the difference between the 2-day average and the long-term average, ensuring the acute and chronic effects are independent. We subtracted the long-term average from the short-term to avoid collinearity issues, and to ensure that differences between ZIP codes in PM$_{2.5}$ at a given time do not contribute to the short-term effect estimate. Hence, the short-term effect cannot be confounded by variables that differ across ZIP codes.

Specifically, we fit a Poisson survival analysis with a logarithmic link function and a log (population) offset term and modeled the expected daily death counts ($\mu_{it}$) in the $i$th ZIP code on the $t$th day as follows:
\log(\mu_{it}) = \lambda_i + \beta_1 PM_{it} + \beta_2 \Delta PM_{it} + \lambda(t) + \text{temporal covariates} + \text{spatial covariates} + \text{offset} \tag{1}

Where \( \lambda_i \) is a random intercept for each ZIP code, \( PM_{it} \) is the 365-day moving average ending on day \( t \) in ZIP code \( i \), \( \Delta PM_{it} \) is the deviation of the 2-day average from its long-term average (\( PM_{it} \)) in ZIP code \( i \), \( \lambda(t) \) is a smooth function of time, temporal covariates are temperature and day of the week, spatial covariates are socio-economic factors defined at ZIP code level (percent of people without high school education, percent of white people, median household income) and smoking data at the county level. Additionally, a Quasipoisson model was used to control for possible overdispersion (Ver Hoef and Boveng 2007).

\( \lambda(t) \) was estimated with a natural cubic spline with 5 degrees of freedom (df) per year, to control for time and season trends. The specific temporal and spatial covariates we used were: a natural cubic spline for temperature with 3 df in total; a categorical variable for day of the week; linear variables for percent of people without high school education, percent of white people, median household income, and percent of people who currently smoke every day.

The number of deaths per ZIP code area over the study period (2003-2008) averaged 319 with a standard deviation of 430. Since the outcome is counts, we cannot adjust for age and sex as in a Cox model. Instead we adjust for variables that vary by ZIP code. The analyses were also repeated without mutual adjustment for short- and long-term PM\(_{2.5}\).

We modeled the association between all-cause mortality and PM\(_{2.5}\) at low doses in which the person-time at risk in each year of follow-up in each ZIP code was used as the
offsets. We also conducted effect modification by population size, by choosing the median (4,628) of ZIP code level total population as the cutoff between urban and rural areas.

**Estimating the effects of low-level PM$_{2.5}$:** For full cohort analyses with 10,938,852 person-years of follow-up, all observed deaths were used. To estimate effects at low levels of exposure we carried out restricted analyses; we conducted one analysis restricted to annual exposures below 10 µg/m$^3$, below the current annual PM$_{2.5}$ NAAQS of 12 µg/m$^3$; and another restricted to observations with short-term exposure below 30 µg/m, below the current daily PM$_{2.5}$ NAAQS of 35 µg/m$^3$. After these exclusions, the chronic analyses were restricted to 268,050 deaths out of 551,024 deaths in total, and the acute analyses to 422,637 deaths.

**Assessing the dose-response relationship:** For both the acute and chronic analyses, we fit penalized regression splines in the restricted analyses to estimate the shape of the dose-response below current EPA standards. The degrees of freedom of the penalized splines for PM$_{2.5}$ was estimated by generalized cross validation (GCV).

**RESULTS**

Table 1 presents a summary of the predicted exposures for both the short-and long-term PM$_{2.5}$ exposure across all grid cells in the study area.

**Table 1.** Descriptive Statistics for PM$_{2.5}$ exposure and temperature in New England, 2003-2008
Table 2 presents the estimated percent change in all-cause mortality with 95% CIs for a 10 µg/m³ increase for both short- and long-term PM$_{2.5}$ in the restricted and full cohort. For the restricted population, we found an estimated 9.28% increase in mortality (95% CI: 0.76, 18.52%) for every 10 µg/m³ increase in long-term PM$_{2.5}$ exposure. For every 10 µg/m³ increase in short-term PM$_{2.5}$ exposure, a 2.14% increase in mortality (95% CI: 1.34, 2.95%) was observed. For long-term exposure, the effect estimates were smaller when higher pollution days were included (7.52%; 95% CI: 1.95, 13.40%), suggesting larger slopes between low-concentration long-term PM$_{2.5}$ and mortality.

### Table 2. Percent increase in mortality (95% CI) for a 10 µg/m³ increase for both short-term and long-term PM$_{2.5}$

<table>
<thead>
<tr>
<th>PM$_{2.5}$ exposure</th>
<th>Model</th>
<th>Percent increase</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>With mutual adjustment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short term PM$_{2.5}$</td>
<td>Low daily exposure $^a$</td>
<td>2.14 ± 0.81</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
### Without mutual adjustment

<table>
<thead>
<tr>
<th></th>
<th>Full cohort</th>
<th>2.14 ± 0.75</th>
<th>&lt;.001</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Long term PM$_{2.5}$</strong></td>
<td>Low chronic exposure $^b$</td>
<td>9.28 ± 8.88</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>Full cohort</td>
<td>7.52 ± 5.73</td>
<td>0.007</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Full cohort</th>
<th>2.07 ± 0.80</th>
<th>&lt;.001</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Short term PM$_{2.5}$</strong></td>
<td>Low daily exposure $^a$</td>
<td>2.08 ± 0.76</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Full cohort</td>
<td>2.08 ± 0.76</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Full cohort</th>
<th>7.16 ± 8.75</th>
<th>0.109</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Long term PM$_{2.5}$</strong></td>
<td>Low chronic exposure $^b$</td>
<td>6.46 ± 5.69</td>
<td>0.026</td>
</tr>
</tbody>
</table>

$^a$The analysis was restricted only to person time with daily PM$_{2.5}$ less than 30 µg/m$^3$ (422,637 deaths).

$^b$The analysis was restricted only to person time with chronic PM$_{2.5}$ less than 10 µg/m$^3$ (268,050 deaths)

The full cohort analysis had 551,024 deaths

Without mutual adjustment, lower estimates were found for both acute and chronic effects, compared to those with mutual adjustment. In full cohort analyses, 2.08% (95% CI: 1.32, 2.84%) and 6.46% (95% CI: 0.93, 12.30%) increase in mortality was found for each 10 µg/m$^3$ increase in short- and long-term PM$_{2.5}$, respectively. In restricted analyses, the corresponding effect estimates were 2.07% (95% CI: 1.27, 2.89%) and 7.16% (95% CI: -1.23, 16.27%), respectively.

Our results were robust to the choice of lag period for acute exposure. We analyzed other averaging periods (Figure 2): for example, lag0 (day of death exposure) and lag04 (a moving average of day of death exposure and previous 4-day exposure). For the acute
effects, we found a significant but smaller association for lag0 PM$_{2.5}$ (1.71%; 95% CI: 1.09, 2.34%) or lag04 PM$_{2.5}$ (1.76%; 95% CI: 0.72, 2.81%) compared to lag01 analysis. The lag period used for short-term exposure did not affect estimates of chronic effects. For example, estimated increases in mortality with a 10 µg/m$^3$ increase in long-term PM$_{2.5}$ were 7.35% (95% CI: 1.79, 13.21%) and 7.25% (95% CI: 1.69, 13.12%) when short-term PM$_{2.5}$ was classified using lag0 or lag04, respectively.

![Figure 2](image)

**Figure 2.** Percent change in mortality per 10 µg/m$^3$ increase in short-term PM$_{2.5}$ with different lags with mutual adjustment. Error bars indicate the 95% CIs.

We also examined effect modification by population size. In the full cohort, a significant interaction for chronic effects was found ($p < 0.01$), with a higher effect of 12.56% (95% CI: 5.71, 19.85%) in urban areas compared to 3.21% (95% CI: -2.92, 9.72%) in rural areas. Such a significant interaction, however, was not observed in the
restricted analysis \((p = 0.16)\). Estimates were 14.27\% (95\% CI: 3.19, 26.53\%) and 5.48\% (95\% CI: -4.21, 16.16\%) in urban and rural areas, respectively. For the short-term exposure, population size did not modify the acute effects neither in full cohort nor in restricted analyses \((p = 0.74\) and 0.46, respectively).

In our penalized spline model for long-term exposure below the cutoff of 10 \(\mu g/m^3\) (Figure 3a), we found a non-linear relationship between long-term PM\(_{2.5}\) and mortality. The association was linear with evidence of a lower slope below 6 \(\mu g/m^3\). However a large confidence interval was observed; hence we could not be confident whether the slope of the dose-response curve for long-term exposures below 6 \(\mu g/m^3\) changed. When examining the shape of the dose-response for chronic effects, both a linear term for short-term exposure \((the\ difference)\) and a penalized spline for long-term average exposure were included in the model, resulting in a penalized spline with a df of 1.71. In contrast, we only included the 2-day average in the penalized spline model of acute effects, to provide a more interpretable dose-response relationship (Figure 3b). Results of this analysis indicated a linear association across the exposure distribution, but we could not be certain about the shape of slope for acute effects below 3 \(\mu g/m^3\).
Figure 3. The dose-response relationship between long-term PM$_{2.5}$ and mortality at low doses with mutual adjustment (a) and the dose-response relationship between short-term PM$_{2.5}$ and mortality at low doses without mutual adjustment (b). Shaded areas indicate the 95% CIs.
DISCUSSION

Applying the predicted daily PM$_{2.5}$ with 1 km spatial resolution from our novel hybrid models, we observed that both short- and long-term PM$_{2.5}$ exposure were significantly associated with all-cause mortality among residents aged 65 and above in New England, even when restricted to ZIP codes and times with annual exposures below 10 μg/m$^3$ or with daily exposure below 30 μg/m$^3$. Hence the association of particle exposure with mortality exists for concentrations below current US, WHO (10 μg/m$^3$ of annual average PM$_{2.5}$, 25 μg/m$^3$ daily) or EU (25 μg/m$^3$ of annual average PM$_{2.5}$) standards (WHO; EU). Notably, this analysis includes all areas in New England and all Medicare enrollees aged 65 and above, and provides chronic effect estimates that are independent of acute effects. Based on a penalized spline model, the positive dose-response relationship between chronic exposure and mortality appears to be linear for PM$_{2.5}$ concentrations down to 6 μg/m$^3$, with a positive (though weaker and less precise) dose-response slope continuing below this level. This lack of power is likely due to the small population exposed in the areas with annual PM$_{2.5}$ below 6 μg/m$^3$, which were quite rural.

For the acute effect, we found a 2.14% (95% CI: 1.38 to 2.89%) increase in all-cause mortality per 10 μg/m$^3$ increment in PM$_{2.5}$ for the full cohort of our study, which is higher than the effect size of most studies using city averages from monitors. For instance, in a US national study by Zanobetti and Schwartz (Zanobetti and Schwartz 2009), the effect size was found to be 0.98% (95% CI: 0.75, 1.22%). Similar results have also been found in a review study, where researchers have demonstrated the overall
summary estimate was 1.04% (95% CI: 0.52, 1.56%) per 10 μg/m³ increment in PM$_{2.5}$ (Atkinson et al. 2014). Most previous studies had exposure data with lower spatial resolution (citywide average, not ZIP code), which introduces exposure measurement error and likely results in downward bias in estimates; these results (for the acute effect) are consistent with such a phenomena. Our restricted study estimated a 2.14% (95% CI: 1.34, 2.95%) increase in all-cause mortality per 10 μg/m³ increment in PM$_{2.5}$, close to the effect size of the full cohort study, possibly because the sample size of the restricted study for acute effect is close to that of the full cohort. That is, the EPA daily standard (35 μg/m³) was almost never exceeded in this study. In addition, lower effect estimates for short-term exposure were observed with mutual adjustment for both full cohort and restricted analyses. This finding has important implications for the interpretation of previous studies without such mutual adjustment.

For the chronic effect, the effect estimate in our full cohort study is consistent with previous studies with comparable sample sizes (Lepeule et al. 2012; Laden et al. 2006; Hoek et al. 2013). For example, ACS study comprised of 500,000 adults from 51 US cities, reported a 6% (95% CI: 2, 11%) increase in all-cause mortality for 10 μg/m³ increment in PM$_{2.5}$ (Pope III et al. 2002). A study of 13.2 million elderly Medicare recipients across the east USA found a 6.8% (95% CI: 4.9, 8.7%) increase in all-cause mortality for 10 μg/m³ increment in PM$_{2.5}$ (Zeger et al. 2008). When we restricted our analysis to annual concentrations below 10 μg/m³, a larger slope of 9.28% (95% CI: 0.76, 18.52%) increase per 10 μg/m³ was observed. Our findings suggest a steeper slope at low concentrations among those aged 65 years and older. This may also reflect particle composition. The sources and composition of the particles may differ on lower pollution
days from that seen on high pollution days, which are probably more impacted by secondary aerosols. Compared to the full cohort, the effect estimate of the restricted analysis was closer to the estimates in the published literature reporting higher slopes, such as the ESCAPE study, the Six City study and Women’s Health Initiative study (Beelen et al. 2014; Puett et al. 2008). Lower effect estimates were also observed for chronic effect without mutual adjustment.

To the best of our knowledge, this study is the first of its kind restricting the exposure and exploring the dose-response relationship between PM$_{2.5}$ and mortality below the current EPA standards (12 μg/m$^3$ of annual average PM$_{2.5}$, 35 μg/m$^3$ daily). Moreover, the use of the MEDICARE cohort means we are studying the entire population of Medicare enrollees over 65, and not a convenience sample. In addition, temperature was also controlled on the 1 × 1 km fine geographic scale. The acute and chronic effects for analyses restricted to low exposure of PM$_{2.5}$ are similar or even higher compared to those of the full cohort analyses. These results indicate that the adverse health effects of PM$_{2.5}$ are at least retained, if not strengthened, at low levels of exposure. However, the findings from the penalized spline model do not support a stronger association at the lowest range of PM$_{2.5}$ concentrations. This finding provides epidemiological evidence for the reevaluation of EPA guidelines and standards, though more evidence are needed to confirm the association below 6 μg/m$^3$.

The Poisson survival analysis applied in this study provides new opportunities in assessing acute and chronic effects simultaneously. As shown in our analysis, the chronic effect estimate was much larger than the acute effect estimate, after controlling for the
acute estimate, indicating there was chronic effect of PM$_{2.5}$, which cannot be solely explained by the short-term exposure.

Another key component of this study is that the application of high spatial- (1 × 1 km) and temporal- (daily) resolution of PM$_{2.5}$ concentrations to some extents reduced exposure error. The out-of-sample $R^2$ is higher than that for the predictions with $10 \times 10$ km spatial resolution.

A potential limitation is the limited availability of individual level confounders, such as smoking status, which could bias the health effect estimates. We were able to control for ZIP code level education, median income, race and county level smoking data. However, a study by Brochu et al. reported that census tract level socioeconomic indicators were uncorrelated with PM$_{2.5}$ on the subregional and local scale, providing some assurance that confounding by SES may not be much of an issue (Brochu et al. 2011). These results suggest that those variables may not confound the association, but the inability to control for them remains an issue. Another limitation is that we did not examine other pollutants such as O$_3$ or NO$_2$ due to lack of data at the same spatial level as PM$_{2.5}$.

**CONCLUSION**

In conclusion, the acute and chronic effects of low-concentration PM$_{2.5}$ were examined for Medicare population using a comprehensive exposure dataset from a satellite-based prediction model. Our findings show that both short- and long-term exposure to PM$_{2.5}$ were associated with all-cause mortality, even for exposure levels not exceeding the newly revised EPA standards, suggesting that adverse health effects occur
at low levels of fine particles. The policy implication is that improving the air quality below the current EPA standards can still yield health benefit.
REFERENCES


