





ARTICLE



<https://doi.org/10.1038/s41467-021-27484-1>

OPEN

Secondary organic aerosol association with cardiorespiratory disease mortality in the United States

Havala O. T. Pye ^{1✉}, Cavin K. Ward-Caviness ², Ben N. Murphy ¹, K. Wyatt Appel¹ & Karl M. Seltzer ³

Fine particle pollution, $PM_{2.5}$, is associated with increased risk of death from cardiorespiratory diseases. A multidecadal shift in the United States (U.S.) $PM_{2.5}$ composition towards organic aerosol as well as advances in predictive algorithms for secondary organic aerosol (SOA) allows for novel examinations of the role of $PM_{2.5}$ components on mortality. Here we show SOA is strongly associated with county-level cardiorespiratory death rates in the U.S. independent of the total $PM_{2.5}$ mass association with the largest associations located in the southeastern U.S. Compared to $PM_{2.5}$, county-level variability in SOA across the U.S. is associated with 3.5× greater per capita county-level cardiorespiratory mortality. On a per mass basis, SOA is associated with a 6.5× higher rate of mortality than $PM_{2.5}$, and biogenic and anthropogenic carbon sources both play a role in the overall SOA association with mortality. Our results suggest reducing the health impacts of $PM_{2.5}$ requires consideration of SOA.

¹Office of Research and Development, U.S. Environmental Protection Agency, 109 TW Alexander Dr, Research Triangle Park, NC 27711, USA. ²Office of Research and Development, U.S. Environmental Protection Agency, 104 Mason Farm Rd, Chapel Hill, NC 27514, USA. ³Oak Ridge Institute for Science and Education Postdoctoral Fellow in the Office of Research and Development, U.S. Environmental Protection Agency, 109 TW Alexander Dr, Research Triangle Park, NC 27711, USA. ✉email: pye.havala@epa.gov

Ambient fine particulate matter, PM_{2.5}, is estimated to be responsible for up to 9 million deaths per year worldwide^{1,2}, including 15 deaths per 100,000 people in the U.S.³. Multiple studies show a link between exposure to PM_{2.5} and cardiovascular and respiratory (cardiorespiratory, CR) disease, hospital admissions, and mortality^{4–8}. While less than ten components of PM_{2.5} contribute significantly to mass, PM_{2.5} includes about 50 easily identifiable inorganic species⁹ and potentially hundreds of thousands of individual organic compounds¹⁰. Total PM_{2.5} mass has been more consistently associated with adverse health outcomes than individual components¹¹, potentially due to limited access to composition in historical epidemiological analyses.

Information on the role of specific sources of organic aerosol (OA) in adverse health outcomes is even more limited than for other components of PM_{2.5}. Recent observations indicate that OA can exceed the mass of other common PM_{2.5} components such as sulfate¹². Secondary organic aerosol (SOA) is produced from reactions of anthropogenic and natural emissions throughout the year^{13,14} and is a dominant component of total OA even in urban locations^{15–17}. While ambient PM_{2.5} concentrations are expected to decline in the future as anthropogenic emissions are controlled, adverse health effects persist below current U.S. air quality standards¹⁸, and a relative increase in SOA could lead to increased health impacts per unit mass¹⁹.

In the southeastern U.S., emissions of monoterpenes and isoprene from vegetation mix with anthropogenic pollutants like nitrogen oxides (NO_x) and other volatile organic compounds (VOCs) to form SOA in high concentration^{13,20–22}. Underlying medical conditions like heart disease as well as CR disease mortality are also higher in the Southeast than the rest of the U.S. as a result of multiple socioeconomic and behavioral factors²³. In addition, higher rates of stroke mortality in the stroke belt²⁴ have persisted for decades²⁵, and cardiovascular mortality rates have been slower to decline in the Southeast than in other regions of the U.S.²³. While previous air quality models struggled to reproduce measured organic carbon (OC), advances in algorithms for SOA formation pathways (oxidation of monoterpenes^{13,22}, isoprene^{20,26}, and anthropogenic volatile organic compounds²⁷) have essentially eliminated OC underestimates in the current Community Multiscale Air Quality (CMAQ) model^{28,29}. Improved SOA estimates combined with multi-decadal reductions of particulate sulfate concentrations allow for an examination of the role of SOA in CR mortality not previously possible. To the best of our knowledge, no epidemiological studies have considered the association of SOA, as estimated by any of the current generation of predictive algorithms, with health outcomes.

In this work, we use year 2016 age-adjusted CR disease mortality rate data from the Centers for Disease Control and Prevention (CDC) and predicted PM_{2.5} in a cross-sectional framework to associate county-level per capita mortality with ambient PM_{2.5} and its components while adjusting for a broad array of relevant confounders (see Methods section). We use multiple linear regression to examine both the contiguous United States as well as the southeastern U.S. for different granularity in OA composition information.

Results

In 2016, CR disease in the U.S. was responsible for a county-level median of 320 age-adjusted deaths per 100,000 in population (2708 counties). The southeastern U.S. experienced a slightly higher rate of 360 age-adjusted deaths per 100,000 in population (646 counties) (Fig. 1a). The predicted county-wide, annual-average U.S. PM_{2.5} concentration was 6.5 μg m⁻³ with OA the

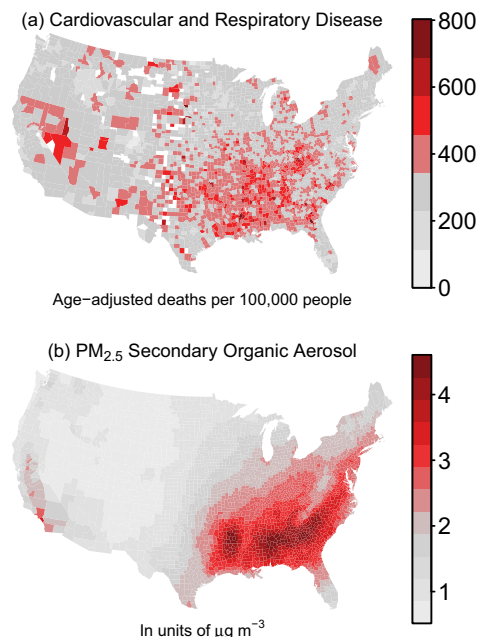


Fig. 1 Cardiorespiratory disease mortality rates and secondary organic aerosol concentrations. County-level, year 2016 (a) cardiovascular and respiratory disease age-adjusted death rates (per 100,000 in population) are from CDC and (b) PM_{2.5} secondary organic aerosol concentrations are predicted by CMAQ. White in (a) indicates no death rate data while light gray indicates low reported rates.

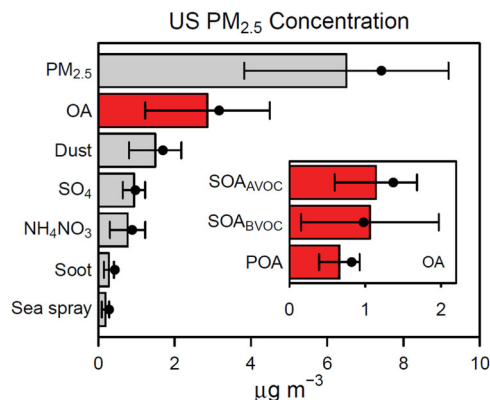


Fig. 2 Contiguous U.S. county-level average predicted concentration of PM_{2.5} and its major components (bars) for 2016 ($n = 2708$ counties).

Major components include organic aerosol (OA), a calcium-iron-silicon-aluminum rich aerosol (dust), sulfate (SO₄), ammonium and nitrate (NH₄NO₃), an elemental carbon and potassium rich aerosol (soot), and a chloride-sodium-magnesium rich aerosol (sea spray). Inset in red are the subcomponents of OA: SOA from anthropogenic VOCs (SOA_{AVOC}), SOA from biogenic VOCs (SOA_{BVOC}), and primary organic aerosol (POA). Error bars represent ±1 IQR variation in the pollutant from the county-level average. Points represent population-weighted concentrations.

most abundant major component at 2.9 μg m⁻³. Other major PM_{2.5} components included a calcium (Ca)-iron (Fe)-silicon (Si)-aluminum (Al)-rich dust, sulfate (SO₄), ammonium and nitrate (NH₄NO₃), elemental carbon- and potassium (K)-rich soot, and chloride (Cl)-sodium (Na)-magnesium (Mg)-rich sea spray aerosol (Fig. 2, see also the Supplementary Information).

PM_{2.5} SOA association with mortality. PM_{2.5}, as well as multiple PM_{2.5} components, were strongly associated with CR mortality

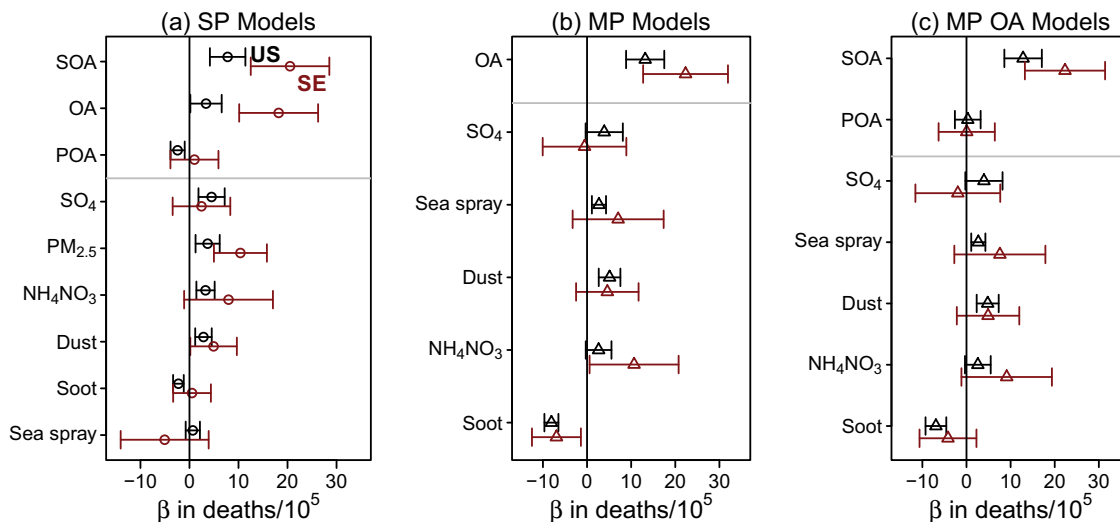


Fig. 3 Association of PM_{2.5} and its components with death rates across the contiguous U.S. (*n* = 2708 counties) in black and southeastern (SE) U.S. (*n* = 646 counties) in red (open symbol) determined via regressed coefficients (β) from multiple linear regression and their 95% confidence intervals (whiskers). Models forms (see Methods, Table 2) are **a single pollutant (SP, circles) or multipollutant (triangles) **(b)** for PM_{2.5} components (MP) and **c** with refinement of OA subcomponents (MP OA). Regressed coefficients correspond to IQR-normalized species concentrations (Supplementary Table 1) in units of deaths per 100,000 in population. Horizontal gray lines are used to visually separate results for OA and its subcomponents (primary organic aerosol, POA, and secondary organic aerosol, SOA) from the other components.**

Table 1 Regression model coefficients (β) and their 95% confidence intervals on a $1 \mu\text{g m}^{-3}$ and IQR-normalized basis for total PM_{2.5} (SP model), OA (MP model), and SOA (MP OA model) for the contiguous U.S. and southeastern U.S. Concentration is population weighted at the county level.

	β [95% CI] for $1 \mu\text{g m}^{-3}$ change in pollutant (deaths $10^{-5} \mu\text{g}^{-1} \text{m}^3$)	β [95% CI] for IQR change in pollutant (deaths 10^{-5})	Average concentration ($\mu\text{g m}^{-3}$)	IQR ($\mu\text{g m}^{-3}$)
<i>Contiguous U.S.</i>				
PM _{2.5}	1.4 [0.5, 2.3]	3.7 [1.2, 6.2]	7.4	2.7
OA	8.1 [5.4, 11]	13 [8.8, 18]	3.2	1.6
SOA	8.9 [6.0, 12]	13 [8.6, 17]	2.3	1.4
<i>Southeastern U.S.</i>				
PM _{2.5}	9.1 [4.3, 14]	10 [5.0, 16]	7.9	1.1
OA	23 [13, 33]	22 [13, 32]	3.8	1.0
SOA	27 [16, 38]	22 [13, 31]	3.0	0.8

across the contiguous U.S. in 2016 (Fig. 3a). Variability in total PM_{2.5} (estimated by the inter-quartile range (IQR)) was associated with an increase of 3.7 (95% CI: 1.2–6.2) CR deaths per 100,000 people (Table 1), a 1.2% increase over the median CR death rate observed. For the southeastern U.S., an IQR increase in PM_{2.5} was associated with 9.1 (95% CI: 4.3–14) additional deaths per 100,000 people or a 2.8% increase over the median observed CR mortality. Associations of the CR death rate with IQR increases in the six primary PM_{2.5} components ranged from –2.3 to 4.5 across the contiguous U.S. and from –5.1 to 18 per 100,000 people across the southeastern U.S. in single pollutant models examining each component in a separate regression model with identical confounder adjustment (see Methods section and Supplementary Table 6).

An IQR increase in the OA component of PM_{2.5} in $\mu\text{g m}^{-3}$ (see Supplementary Table 1 for IQR values) was more strongly associated with CR mortality than was total PM_{2.5}. The initial

multipollutant (MP) model with six primary PM_{2.5} components indicated OA had the largest association by magnitude ($\beta = 13$, 95% CI: 8.8–18, deaths per 100,000) for an IQR change in concentration (Fig. 3b). Thus, an IQR increase in OA was associated with a greater elevation in county-level CR death rates than an IQR change in sulfate, ammonium and nitrate, sea spray, dust, or soot. When OA was further divided into primary organic aerosol (POA) and SOA (Fig. 3c), SOA ($\beta = 13$, 95% CI: 8.6–17, deaths per 100,000) but not POA ($\beta = 0.29$, 95% CI: –2.6–3.2, deaths per 100,000) was strongly associated with CR mortality. Similar results were obtained for the southeastern U.S. with SOA ($\beta = 22$, 95% CI: 13–31, deaths per 100,000 people) being more strongly associated with CR mortality than any other component and having a stronger association than observed in the entire U.S. The number of per capita deaths associated with IQR variability in SOA was 2.1 and 3.5 times larger than the number associated with total PM_{2.5} in the Southeast and contiguous U.S., respectively.

Role of SOA beyond PM_{2.5} mass. To determine whether the association between SOA and CR mortality was independent of PM_{2.5} mass, additional models were adjusted for PM_{2.5} mass (SP-adj) or used residuals (R) resulting from the regression of SOA on PM_{2.5} mass as an exposure³⁰ (see Methods section and Table 2). In the SP-adj framework, IQR variability in SOA was positively associated with mortality and had a larger regression coefficient than sulfate, ammonium and nitrate, sea spray, dust, soot, and POA as well as OA over both the contiguous and southeastern U.S. This indicates the association between SOA and CR mortality was not primarily driven by changes in total PM_{2.5}. The SOA component residual was also associated with increased CR mortality rates across the U.S. and Southeast (Supplementary Fig. 4).

SOA was also associated with higher CR mortality rates on a $1 \mu\text{g m}^{-3}$ mass basis. Across the contiguous U.S., SOA was associated with 8.9 (95% CI: 6.0–12) additional deaths per 100,000 in population for a $1 \mu\text{g m}^{-3}$ increase in concentration indicating about 6.5 × higher association than total PM_{2.5} (see Table 1). In the southeastern U.S., SOA was associated with 27

Table 2 Regression model forms.

Model	Form	Description
Single Pollutant (SP)	$D = \beta_0 + \beta_1 PM_j + \sum_{i=1}^N \beta_i C_i$	Primary models evaluating association of PM _{2.5} or a component with mortality
Single Pollutant-Adjusted for total PM _{2.5} (SP-adj)	$D = \beta_0 + \beta_1 PM_j + \beta_2 PM_{2.5} + \sum_{i=1}^N \beta_i C_i$	SP model with adjustment of components for total PM _{2.5}
Residual (R)	$D = \beta_0 + \beta_1 residual_i + \sum_{i=1}^N \beta_i C_i$ where $residual_i = PM_i - m_i \times PM_{2.5} - b_i$ and m_i and b_i are component-specific coefficients obtained from the regression: $PM_i = m_i \times PM_{2.5} + b_i$	Second adjustment model of PM _{2.5} components for total PM _{2.5} based on residuals from regressing total PM _{2.5} components on PM _{2.5}
Multipollutant (MP)	$D = \beta_0 + \{\beta_1 PM_{OA}\} + \beta_2 PM_{Sespray} + \beta_3 PM_{Soot} + \beta_4 PM_{NH_3} + \beta_5 PM_{SO_2} + \beta_6 PM_{Dust} + \sum_{i=1}^N \beta_i C_i$	Model with 6 primary PM _{2.5} components including OA to evaluate their independent associations
Multipollutant OA (MP OA)	$D = \beta_0 + \{\beta_1 PM_{SOA} + \beta_2 PM_{POA}\} + \beta_3 PM_{Sespray} + \beta_4 PM_{Soot} + \beta_5 PM_{NH_3} + \beta_6 PM_{SO_2} + \beta_7 PM_{Dust} + \sum_{i=1}^N \beta_i C_i$	Multipollutant composition model as in MP but with refinement of OA (OA subdivided into SOA and POA)
Multipollutant SOA (MP SOA)	$D = \beta_0 + \{\beta_1 PM_{SOA} + \beta_2 PM_{POA}\} + \beta_3 PM_{Sespray} + \beta_4 PM_{Soot} + \beta_5 PM_{NH_3} + \beta_6 PM_{SO_2} + \beta_7 PM_{Dust} + \sum_{i=1}^N \beta_i C_i$	Multipollutant composition model as in MP OA but with refinement of SOA into SOA from oxidation of VOCs that are predominantly biogenic (BVOC) and anthropogenic (AVOC) in origin.

D, the outcome, is the age-adjusted cardiovascular and respiratory disease death rate in number of deaths per 100,000 people. By default, PM_i are concentrations of component *i* in µg m⁻³ normalized to the IQR, also in µg m⁻³, of that pollutant. When coefficients are presented per unit mass, the IQR normalization does not occur. PM_{2.5} is total PM_{2.5} concentration; β values indicate regressed coefficients. Confounders (N = 28 by default) with values, C_i, are included in each model. Bold indicates a vector of county-level values. In MP models, brackets are placed around OA components for emphasis.

(95% CI: 16–38) additional deaths per 100,000 in population for a 1 µg m⁻³ increase in concentration or about a 3.0× higher association than PM_{2.5}. A 1 µg m⁻³ change in SOA in the southeastern U.S. also showed a 3.0× higher association with per capita CR mortality than for nationwide conditions. For 1 µg m⁻³ of PM_{2.5}, the association was 6.6× higher in the Southeast than across the U.S.

Role of SOA subcomponents. A multipollutant model (MP SOA) was used to estimate the health effects of SOA sub-components including SOA from VOCs that were predominantly biogenic in origin (SOA_{BVOC}) versus those that were anthropogenic in origin (SOA_{AVOC}). SOA_{BVOC} showed a positive association with death rates across the U.S. (β = 10, 95% CI: 5.8–14, deaths per 100,000 people for the contiguous U.S. per IQR change, IQR = 0.90 µg m⁻³) as well as in the southeastern U.S. (β = 25, 95% CI: 16–33, deaths per 100,000 people per IQR change; IQR = 0.72 µg m⁻³) (Fig. 4, MP SOA). Sensitivity models on both the contiguous and southeastern U.S. domains showed SOA from biogenic VOCs was associated with CR mortality even after adjustment for total PM_{2.5}, although the SOA_{BVOC}-CR mortality association depended on the PM_{2.5} adjustment method (Supplementary Fig. 4). Further subdivision of SOA_{BVOC} into subcomponents representing the SOA from oxidation of terpenes and isoprene also showed positive associations of each sub-component with CR mortality; however, relationships varied among model frameworks (Supplementary Figs. 4 and 5), and terpene- versus isoprene-derived SOA may be difficult to robustly separate due to correlation between them on a nationwide basis (r = 0.79).

Anthropogenic-VOC derived SOA was slightly more abundant than SOA_{BVOC} both nationwide (Fig. 2, inset) and in the southeastern U.S. with locations near Los Angeles, Atlanta, New York, and other cities showing local maximums in concentration. Multipollutant model results indicated higher levels of anthropogenic SOA were associated with higher county-level CR mortality rates across the contiguous U.S. (β = 3.4, 95% CI: 0.29–6.4, deaths per 100,000 people per IQR change) but indicated an inverse (β = -11, 95% CI: -21 to -1.4, deaths per 100,000 people per IQR change) association with death rates for the southeastern U.S. (Fig. 4, MP SOA). Sensitivity modeling showed the region of interest (contiguous vs southeastern U.S.) affected the direction of association for SOA_{AVOC}.

Utility of observed OC. OC is the only routinely measured indicator of organic PM_{2.5} at EPA Air Quality System (AQS) sites. Predicted OC was highly correlated with OA and SOA from the model (Pearson r = 0.97 or higher). Like OA, increases in model-predicted OC for the contiguous U.S. (2708 counties), showed a positive association with CR mortality (β = 2.3, 95% CI: -1.1–5.6, deaths per 100,000 people per IQR change, IQR = 0.92 µg m⁻³). After considering the role of PM_{2.5} mass (SP-adj or R model framework), the association of model-predicted OC with CR mortality remained positive for the southeastern U.S. only. Sub-setting the complete contiguous U.S. model-predicted OC (n = 2708) to the AQS sites with data for 2016 (n = 232), resulted in no association between predicted OC and increased CR mortality rates (β = -1.3, 95% CI: -12–9.3, deaths per 100,000 people per IQR change, IQR = 1.2 µg m⁻³). Measured OC at AQS sites indicated an association between observed OC and county-level CR mortality rates across the U.S. (β = 4.1, 95% CI: 1.1–7.0, deaths per 100,000 people per IQR change, IQR = 1.2 µg m⁻³), suggesting, measured OC has greater power (smaller confidence interval with larger magnitude association) than modeled OC in a direct comparison.

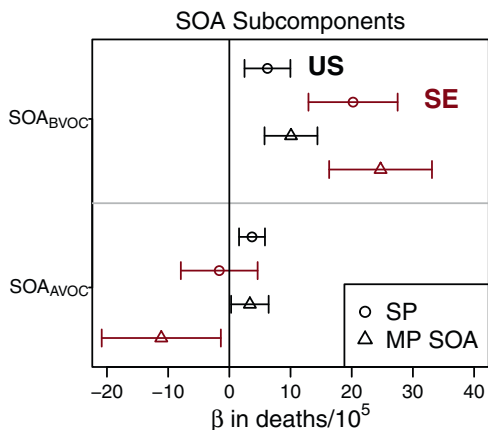


Fig. 4 Association of PM_{2.5} SOA subcomponents with death rates across the contiguous U.S. ($n = 2708$ counties) in black and southeastern (SE) U.S. ($n = 646$ counties) in red (open symbol) determined via regressed coefficients (β) from multiple linear regression and their 95% confidence intervals (whiskers). Model forms are single pollutant (SP, circles) and multipollutant with refinement of SOA subcomponents (MP SOA, triangles). SOA_{BVOC} and SOA_{AVOC} are from the oxidation of VOCs that are predominantly biogenic and anthropogenic in origin, respectively. Regressed coefficients correspond to IQR-normalized species concentrations in units of deaths per 100,000 in population.

Discussion

This work demonstrates a robust association between annual-average SOA concentrations and county-level CR death rates in a cross-sectional study. Associations were elevated in the southeastern U.S. as compared to the contiguous U.S. Not only was SOA associated with higher CR death rates than other components considered, it showed a larger association per unit mass than total PM_{2.5}. The role of SOA was separately distinguished from that of total PM_{2.5} mass as indicated by single pollutant adjusted for PM_{2.5} and component residual model results.

Laboratory experiments indicate a role for SOA in adverse health outcomes. SOA exposure has been shown to result in lung cell death³¹ and formation of reactive oxygen species that can lead to adverse health effects^{32–35}. Acellular assays of laboratory-generated SOA from terpenes, isoprene, and aromatics show oxidative potential that can approach that of ambient PM_{2.5} and known hazardous air pollutants such as diesel exhaust particles^{35–38}. Continued atmospheric processing, including oxidation of SOA or POA also leads to increased oxidative potential^{35,39}. In addition, exposure to isoprene-derived SOA has been shown to modify gene expression in human airway epithelial cells⁴⁰. CR mortality associations with SOA in this work were not driven by a single SOA subcomponent and a consensus on which SOA systems drive redox reactivity as determined by dithiothreitol acellular assays is also lacking³⁴. However, Bates et al.³⁵ indicate the highest oxidative potential for aromatic (predominantly anthropogenic) SOA followed by terpene then isoprene SOA. Combustion sources, like biomass burning and vehicles which lead to both primary and secondary OA, are likely a significant source of oxidative potential in urban areas⁴¹ and have been linked with pediatric respiratory disease emergency visits in the eastern U.S.⁴².

Anthropogenic SOA is more abundant than biogenic SOA and POA in locations like the Northeast and California (Fig. 2 and Supplementary Fig. 1j–l), but it did not have a higher association with CR mortality when compared to other SOA subcomponents and showed no association when analysis was restricted to the southeastern U.S. Anthropogenic SOA model algorithms are

generally more empirical than for terpene and isoprene SOA pathways. In addition, significantly less data is available from urban atmospheres in the U.S. that could be used to better elucidate chemical pathways of anthropogenic SOA. Therefore, the mixed associations for anthropogenic SOA on U.S. vs southeastern U.S. domains should be revisited as model representations continue to improve or as observational data sets for similar analysis become available. Biogenic SOA was robustly associated with higher CR mortality. While the ability to separate biogenic SOA into further subcomponents may be limited due to correlation, where correlation between the isoprene and monoterpene SOA was lowest (in the Southeast, $r = 0.54$) both pollutants showed positive associations with CR mortality.

Future work should continue to quantify the effects of different PM components on health, with consideration of SOA and its subcomponents. Historically, the association of individual PM_{2.5} subcomponents and health effects have lacked consistency, and as a result, regulation of PM_{2.5} in the U.S. has focused on total mass¹¹. SOA concentration measurements are not routinely available, so the analysis shown here relied upon the ability of a model to robustly predict SOA. New data sets of measured SOA concentrations could be directly used in health impact assessment and lead to improved air quality models which will also ultimately lead to refinement of health effect estimates. An improved understanding of SOA health implications could influence control strategies to effectively reduce mortality. For example, isoprene SOA is likely best reduced by targeting sulfur oxide emissions²⁰ while monoterpene SOA can be reduced via controls on NO_x^{21,22}. In addition, an improved representation of SOA from anthropogenic sources could improve health impact estimates and mitigation by providing insight into which sources (vehicle, evaporative chemical products, commercial and residential cooking, solid fuel burning, etc.) most affect public health.

This study benefited from having a large dataset spanning 2708 counties across the U.S. and high-quality air pollutant concentration data. The model algorithms and processes responsible for SOA prediction were previously evaluated against data from field campaigns^{13,26,27,43} as well as nationwide OC observations^{27,29}. This allowed for a systematic examination of PM_{2.5} components, specifically SOA, that is not standardly available. In addition, results from initial statistical models were examined through sensitivity simulations confirming the association between SOA and CR mortality. However, the analysis performed here does not exclude a role for components other than SOA from being important for health. Primary particle effects may be acute and on smaller spatio-temporal scales than examined here. Future work focused on individual health outcomes and exposures, rather than county-level information, could better characterize near-source exposure and consequent health implications, which may be more relevant for some components and communities⁴⁴.

The analysis here using measured OC is consistent with SOA showing strong association with CR mortality. The ability of measured OC to show a more precise relationship with cardiorespiratory mortality than modeled OC suggests that increasing the spatial coverage of OC measurements could be beneficial for understanding the health impacts of PM_{2.5} components, particularly in the southeastern U.S. where the limited availability of AQS sites prohibited an observation-based analysis. Total OC lacks the distinction between POA and SOA which could explain why both positive and negative associations of OC with adverse health endpoints have been found in previous work^{11,45,46}. However, new observations could employ online techniques⁴⁷ to mitigate filter artifacts⁴⁸, and measurement of specific chemical fragments⁴⁷ or functional groups⁴⁹ could lead to a better understanding of the role of OA subcomponents. Multiple studies have shown that combining air quality models and measurements

together will likely provide the most robust information going forward^{18,50}. As new SOA exposure estimates are created for additional years and through other techniques, similar analysis should be performed to further examine the role of SOA.

Health benefits due to changes in fine particle concentrations are quantified for multiple purposes including evaluation of the public health benefits of pollution reduction efforts⁵¹, co-benefits of carbon pricing⁵², effectiveness of emission controls⁵³, and forecasting air quality impacts under future scenarios⁵⁴. In each case, the mortality-related impacts of PM_{2.5} are influenced by changes in the total as dictated by the changes in individual components⁵⁵. SOA is often explicitly excluded in these analyses^{52,56,57}. The work here suggests a focus on primary emissions or inorganic components alone is insufficient to properly estimate health impacts of changes in PM_{2.5}. If recent trends from the past decade continue^{58,59}, sulfate will continue to decrease as a fraction of PM_{2.5} in favor of OA. Since SOA is an important driver of the health impacts of PM_{2.5} examined here, strategies to further reduce health impacts of PM_{2.5} via emission controls, as well as understanding how death rates have changed in the recent past, requires consideration of SOA.

Methods

This cross-sectional study used multiple linear regression to estimate how spatial variability in pollutant concentrations is associated with CR death rates while adjusting for a wide array of confounders. Multiple sensitivity analyses were undertaken to understand the robustness of the associations and independence of estimated health effects.

Pollutant concentrations. County-level pollutant concentrations (full description in Supplementary Table 2) were created by mapping 12 km by 12 km horizontal resolution CMAQ v5.3.1^{28,29} (Supplementary Table 8) predictions for 2016 by grid cell center latitude and longitude to county locations using the R (v3.6.2) libraries `sp`⁶⁰, `maps`⁶¹, and `maptools`⁶². When multiple CMAQ cell centers fell within one county, all respective CMAQ cells were averaged to create the county-level prediction. Four small counties did not contain any CMAQ grid cell center. These predictions were created by identifying the most populous city in each of the four counties⁶³ and the CMAQ grid cell containing that city. CMAQv5.3.1 tends to be relatively unbiased in its prediction of annual-average PM_{2.5} and OC on a nationwide basis (<10% absolute normalized mean bias compared to AQS observations; Supplementary Table 9 and Supplementary Fig. 6). Predictions for specific OA systems have been evaluated in previous work (summarized in Supplementary Table 10)^{13,26,27}, and predictions from v5.3.1 for the summer in the Southeast are greatly improved over previous model versions.

Total OC in PM_{2.5} measured at Chemical Speciation Network (CSN) and the Interagency Monitoring of Protected Visual Environments (IMPROVE) network sites during 2016 was obtained from the EPA AQS. Both CSN and IMPROVE use a thermal optical reflectance method to measure OC collected on filters and report similar results for collocated samplers (12% bias)⁶⁴. Observed OC data were aggregated to an annual average and multiple sites within a county were averaged into one county-level value for pairing with the health outcome. CMAQ OC predictions were determined by converting concentrations of OA by model species to OC using species-specific organic matter to OC ratios^{13,26,27,43}.

Cardiorespiratory deaths. The outcome of interest was year 2016 age-adjusted CR mortality rates as determined by the Centers for Disease Control and Prevention⁶⁵. Age-adjusted rates are weighted by the population age relative to U.S. standard population and created by the CDC⁶⁶. CR mortality was determined via ICD-10 codes I00-I99 (diseases of the circulatory system) and J00-J98 (diseases of the respiratory system). All I00-J98 deaths, rather than specific disease outcomes such as stroke, were used to increase sample size and power, particularly in low-population rural areas where deaths for specific causes can be very low and thus poorly estimated. Counties with missing death rates (CDC does not report rates when the count is below 9) were removed from the analysis. The CDC flagged counties with 20 or fewer CR deaths as having unreliable rates due to increased variance in the numerator. These counties remained in the analysis to increase the nationwide representativeness despite the potential for increased noise. County population was included as a weight in all models to better account for the precision in estimating rates (higher population counties naturally have more precision in estimating mortality rates) and account for the increased noise associated with low-population counties. In addition to the contiguous United States (48 states), subgroup analysis was performed on the southeastern U.S., defined as Kentucky, Tennessee, North Carolina, Mississippi, Alabama, Georgia, South Carolina, and Florida (EPA Region 4, <https://www.epa.gov/aboutepa/visiting-regional-office>).

Confounders. County-level confounders including population demographics (e.g. race, sex, age), behaviors (e.g. smoking, obesity), social and economic factors (e.g. education, income, unemployment, insured adults), and environmental parameters (water quality, relative humidity, temperature) were included in all statistical models. Relative humidity and temperature were obtained from WRF v4.1.1 predictions and processed through the CMAQ modeling system including the Meteorology-Chemistry Interface Processor⁶⁷ as in the work of Appel, et al.²⁹. All other confounders were obtained from the University of Wisconsin Population Health Institute's (UWPHI) County Health Rankings (CHR)⁶⁸ 2018 dataset⁶⁹. From the full set of CHR measures, metrics that were duplicative (e.g. demographic metrics that summed to one), a health outcome (e.g. poor health, life expectancy), or likely irrelevant to the outcome (e.g. driving alone, dentists, teen births) were removed from consideration. In addition, we only considered CHR factors with values for >95% of the contiguous U.S. counties. Combined with relative humidity and temperature, this resulted in 28 confounders. Underlying inputs to the UWPHI data come from the CDC, Bureau of Labor Statistics, American Community Survey, Census Bureau, Department of Agriculture, Centers for Medicare and Medicaid Services as well as other data sources and were generally representative of 2016 information. Supplementary Table 3 provides the full list of confounders along with their original data source and year represented. The statistical model input dataset, including predicted composition, outcome, and confounders, resulted in complete information for 2708 (out of 3075) counties for the contiguous U.S. and 646 (out of 736) counties for the southeastern U.S. The minimum, mean, maximum, and IQR for all confounders are summarized in Supplementary Table 1.

Statistical models. Primary analyses use multiple linear regression to associate PM_{2.5} with CR mortality while adjusting for the confounders described above. County-level population (American Community Survey 5-year estimate) was included as a weight in each model as mentioned previously. In addition to total PM_{2.5}, PM_{2.5} components were analyzed. Individual PM_{2.5} species estimated by CMAQ were grouped into six main components, mutually exclusive of each other and summing to total PM_{2.5}: OA, sea spray, soot, NH₄NO₃, SO₄, and dust. These components were created because many individual PM_{2.5} species are highly correlated (Supplementary Table 4) which limits the ability of statistical models to separate their individual health associations. The six main components had correlations with each other that did not exceed 0.67 (Pearson *r*, Supplementary Table 4). The conversion from raw CMAQ output to model components used in this work (Supplementary Table 2) along with the spatial distribution of concentrations (Supplementary Fig. 1) is available in the supplementary information.

In the second stage of models, the associations of PM_{2.5} components with mortality were adjusted for total PM_{2.5} both by including a term for PM_{2.5} within the models (SP-adj) and by regressing each component on total PM_{2.5} and taking the residual³⁰ for the component variation that was not due to total PM_{2.5} variation (R) (Table 2). Since residual models remove the portion of the pollutant that is correlated with total PM_{2.5}, any statistically significant positive association of the residual with the outcome indicates that variability in that component specifically contributes to the adverse health effect beyond its role in PM_{2.5} total mass.

To further explore the relationship between the components, a series of analyses included multiple components within a single model and primarily focused on whether associations with organic PM_{2.5}, and further subdivisions of organic PM_{2.5}, were independent of other PM_{2.5} components. These multipollutant models were constructed to provide increasing granularity of information on associations between SOA and SOA subcomponents. The first multipollutant composition model (MP) included OA as well as the five other major PM_{2.5} components in a single model. The second composition model (MP OA) further divided OA into POA and SOA. The correlation between PM_{2.5} components including POA and SOA did not exceed *r* = 0.79. The third, composition model (MP SOA) further divided SOA into two subcomponents based on the dominant origin (biogenic vs anthropogenic) of their precursor VOCs. SOA from C₁₀-C₁₅ terpenes as well as isoprene and related (e.g. glyoxal) species was labeled as SOA_{BVOC} while SOA from anthropogenic VOCs was labeled SOA_{AVOC} (Supplementary Table 2). The models are described in Table 2 and Supplementary Fig. 2.

Splines⁷⁰ as implemented in generalized additive models (R package `mgcv`⁷¹) were used to construct concentration-response curves for the association between PM_{2.5} and county-level cardiorespiratory mortality rate (see Supplementary Fig. 3). All statistical models used R v4.0.0⁷². All regressed coefficients are presented per IQR increase in the exposure along with a 95% confidence interval unless otherwise indicated.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The complete set of county-level concentrations and multiple regression results generated in this study have been deposited at <https://doi.org/10.5281/zenodo.5713903>. Mortality rate data used in this work are available from the Centers for Disease Control and Prevention, National Center for Health Statistics, CDC WONDER Online Database, released June 2017. Data are from the Compressed Mortality File 1999–2016 Series 20

No. 2U, 2016, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Last accessed at <http://wonder.cdc.gov/cmf-icd10.html> on 31 Aug 2020 9:15:06 a.m. County Health Rankings 2018 data developed as a collaboration between the Robert Wood Johnson Foundation and the University of Wisconsin Population Health Institute were obtained from <https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation> (last access: 31 March 2020)⁶⁹. Observed air quality data is available from the EPA Air Quality System (AQS) available at <https://www.epa.gov/aqs> (last access: 18 October 2021)^{73,74}. Source data for data generated in this work are provided with this paper. Source data are provided with this paper.

Code availability

The analysis code for post-processing CMAQ predictions, aggregating CMAQ to annual-average component values, aligning data by county, and performing the statistical analyses have been deposited at <https://doi.org/10.5281/zenodo.571390374>. CMAQ v5.3.1 is available at <https://doi.org/10.5281/zenodo.358589828>.

Received: 9 October 2020; Accepted: 19 November 2021;

Published online: 16 December 2021

References

- Cohen, A. J. et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet* **389**, 1907–1918 (2017).
- Burnett, R. et al. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc. Natl Acad. Sci. USA* **115**, 9592 (2018).
- Health Effects Institute. State of Global Air 2019, https://www.stateofglobalair.org/sites/default/files/soga_2019_report.pdf (2019).
- Al-Kindi, S. G., Brook, R. D., Biswal, S. & Rajagopalan, S. Environmental determinants of cardiovascular disease: lessons learned from air pollution. *Nat. Rev. Cardiol.* **17**, 656–672 (2020).
- Zanobetti, A., Franklin, M., Koutrakis, P. & Schwartz, J. Fine particulate air pollution and its components in association with cause-specific emergency admissions. *Environ. Health* **8**, 58 (2009).
- Peng, R. D. et al. Emergency admissions for cardiovascular and respiratory diseases and the chemical composition of fine particle air pollution. *Environ. Health Perspect.* **117**, 957–963 (2009).
- Dominici, F. et al. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *J. Am. Med. Assoc.* **295**, 1127–1134 (2006).
- Pope, C. A. et al. Cardiovascular mortality and long-term exposure to particulate air pollution. *Circulation* **109**, 71–77 (2004).
- Bell, M. L., Dominici, F., Ebisu, K., Zeger, S. L. & Samet, J. M. Spatial and temporal variation in PM_{2.5} chemical composition in the United States for health effects studies. *Environ. Health Perspect.* **115**, 989–995 (2007).
- Goldstein, A. H. & Galbally, I. E. Known and unexplored organic constituents in the Earth's atmosphere. *Environ. Sci. Technol.* **41**, 1514–1521 (2007).
- U.S. Environmental Protection Agency. *Integrated Science Assessment (ISA) for Particulate Matter*. Report No. EPA/600/R-19/188 (2019).
- Attwood, A. R. et al. Trends in sulfate and organic aerosol mass in the Southeast U.S.: Impact on aerosol optical depth and radiative forcing. *Geophys. Res. Lett.* **41**, 7701–7709 (2014).
- Xu, L. et al. Experimental and model estimates of the contributions from biogenic monoterpenes and sesquiterpenes to secondary organic aerosol in the southeastern United States. *Atmos. Chem. Phys.* **18**, 12613–12637 (2018).
- Shah, V. et al. Widespread pollution from secondary sources of organic aerosols during winter in the northeastern United States. *Geophys. Res. Lett.* **46**, 2974–2983 (2019).
- Volkamer, R. et al. Secondary organic aerosol formation from anthropogenic air pollution: Rapid and higher than expected. *Geophys. Res. Lett.* <https://doi.org/10.1029/2006GL026899> (2006).
- Hayes, P. L. et al. Organic aerosol composition and sources in Pasadena, California, during the 2010 CalNex campaign. *J. Geophys. Res. -Atmos.* **118**, 9233–9257 (2013).
- Jimenez, J. L. et al. Evolution of organic aerosols in the atmosphere. *Science* **326**, 1525 (2009).
- Di, Q. et al. Air pollution and mortality in the medicare population. *New Engl. J. Med.* **376**, 2513–2522 (2017).
- Zhang, W. et al. Triggering of cardiovascular hospital admissions by fine particle concentrations in New York state: Before, during, and after implementation of multiple environmental policies and a recession. *Environ. Pollut.* **242**, 1404–1416 (2018).
- Pye, H. O. T. et al. Epoxide pathways improve model predictions of isoprene markers and reveal key role of acidity in aerosol formation. *Environ. Sci. Technol.* **47**, 11056–11064 (2013).
- Pye, H. O. T. et al. Anthropogenic enhancements to production of highly oxygenated molecules from autoxidation. *Proc. Natl Acad. Sci. USA* **116**, 6641 (2019).
- Pye, H. O. T. et al. Modeling the current and future roles of particulate organic nitrates in the southeastern United States. *Environ. Sci. Technol.* **49**, 14195–14203 (2015).
- Singh, G. K., Azuine, R. E., Siahpush, M. & Williams, S. D. Widening geographical disparities in cardiovascular disease mortality in the United States, 1969–2011. *Int. J. MCH AIDS* **3**, 134–149 (2015).
- Lanska, D. J. & Kuller, L. H. The geography of stroke mortality in the United States and the concept of a stroke belt. *Stroke* **26**, 1145–1149 (1995).
- Glymour, M. M., Kosheleva, A. & Boden-Albala, B. Birth and adult residence in the Stroke Belt independently predict stroke mortality. *Neurology* **73**, 1858–1865 (2009).
- Pye, H. O. T. et al. On the implications of aerosol liquid water and phase separation for organic aerosol mass. *Atmos. Chem. Phys.* **17**, 343–369 (2017).
- Murphy, B. N. et al. Semivolatile POA and parameterized total combustion SOA in CMAQv5.2: impacts on source strength and partitioning. *Atmos. Chem. Phys.* **17**, 11107–11133 (2017).
- U.S. EPA Office of Research and Development. CMAQ (Version 5.3.1), <https://doi.org/10.5281/zenodo.3585898> (2019).
- Appel, K. W. et al. The Community Multiscale Air Quality (CMAQ) model versions 5.3 and 5.3.1: system updates and evaluation. *Geosci. Model Dev.* **14**, 2867–2897 (2021).
- Mostofsky, E. et al. Modeling the association between particle constituents of air pollution and health outcomes. *Am. J. Epidemiol.* **176**, 317–326 (2012).
- Baltensperger, U. et al. Combined determination of the chemical composition and of health effects of secondary organic aerosols: The POLYSOA project. *J. Aerosol Med. Pulm. Drug Deliv.* **21**, 145–154 (2008).
- Nel, A. Air pollution-related illness: effects of particles. *Science* **308**, 804 (2005).
- Ayres, J. G. et al. Evaluating the toxicity of airborne particulate matter and nanoparticles by measuring oxidative stress potential—A workshop report and consensus statement. *Inhal. Toxicol.* **20**, 75–99 (2008).
- Shiraiwa, M. et al. Aerosol health effects from molecular to global scales. *Environ. Sci. Technol.* **51**, 13545–13567 (2017).
- Bates, J. T. et al. Review of acellular assays of ambient particulate matter oxidative potential: Methods and relationships with composition, sources, and health effects. *Environ. Sci. Technol.* **53**, 4003–4019 (2019).
- Kramer, A. J. et al. Assessing the oxidative potential of isoprene-derived epoxides and secondary organic aerosol. *Atmos. Environ.* **130**, 211–218 (2016).
- Jiang, H., Jang, M., Sabo-Attwood, T. & Robinson, S. E. Oxidative potential of secondary organic aerosols produced from photooxidation of different hydrocarbons using outdoor chamber under ambient sunlight. *Atmos. Environ.* **131**, 382–389 (2016).
- Tuet, W. Y. et al. Chemical oxidative potential of secondary organic aerosol (SOA) generated from the photooxidation of biogenic and anthropogenic volatile organic compounds. *Atmos. Chem. Phys.* **17**, 839–853 (2017).
- Chowdhury, P. H. et al. Exposure of lung epithelial cells to photochemically aged secondary organic aerosol shows increased toxic effects. *Environ. Sci. Technol. Lett.* **5**, 424–430 (2018).
- Lin, Y.-H. et al. Gene expression profiling in human lung cells exposed to isoprene-derived secondary organic aerosol. *Environ. Sci. Technol.* **51**, 8166–8175 (2017).
- Bates, J. T. et al. Source impact modeling of spatiotemporal trends in PM_{2.5} oxidative potential across the eastern United States. *Atmos. Environ.* **193**, 158–167 (2018).
- Huang, M., Ivey, C., Hu, Y., Holmes, H. A. & Strickland, M. J. Source apportionment of primary and secondary PM_{2.5}: associations with pediatric respiratory disease emergency department visits in the U.S. State of Georgia. *Environ. Int.* **133**, 105167 (2019).
- Qin, M. et al. Criteria pollutant impacts of volatile chemical products informed by near-field modelling. *Nat. Sustain.* **4**, 129–137 (2021).
- Do, K. et al. A data-driven approach for characterizing community scale air pollution exposure disparities in inland Southern California. *J. Aerosol Sci.* **152**, 105704 (2021).
- Ostro, B. et al. Assessing long-term exposure in the California Teachers Study. *Environ. Health Perspect.* **119**, A242–A243 (2011).
- Thurston, G. D. et al. Ischemic heart disease mortality and long-term exposure to source-related components of U.S. fine particle air pollution. *Environ. Health Perspect.* **124**, 785–794 (2016).
- Ng, N. L. et al. An Aerosol Chemical Speciation Monitor (ACSM) for routine monitoring of the composition and mass concentrations of ambient aerosol. *Aerosol Sci. Technol.* **45**, 780–794 (2011).

48. El-Sayed, M. M. H., Amenumey, D. & Hennigan, C. J. Drying-induced evaporation of secondary organic aerosol during summer. *Environ. Sci. Technol.* **50**, 3626–3633 (2016).
49. Reggente, M., Dillner, A. M. & Takahama, S. Analysis of functional groups in atmospheric aerosols by infrared spectroscopy: systematic intercomparison of calibration methods for US measurement network samples. *Atmos. Meas. Tech.* **12**, 2287–2312 (2019).
50. Ward-Caviness, C. K. et al. Associations between long-term fine particulate matter exposure and mortality in heart failure patients. *J. Am. Heart Assoc.* **9**, e012517 (2020).
51. U.S. Environmental Protection Agency. *Control of Air Pollution from Motor Vehicles: Tier 3 Motor Vehicle Emission and Fuel Standards Final Rule: Regulatory Impact Analysis*. Report No. EPA-420-R-14-005, (2014).
52. Li, M. et al. Air quality co-benefits of carbon pricing in China. *Nat. Clim. Change* **8**, 398–403 (2018).
53. Zhang, Y. et al. Long-term trends in the ambient PM_{2.5}- and O₃-related mortality burdens in the United States under emission reductions from 1990 to 2010. *Atmos. Chem. Phys.* **18**, 15003–15016 (2018).
54. Silva, R. A. et al. The effect of future ambient air pollution on human premature mortality to 2100 using output from the ACCMIP model ensemble. *Atmos. Chem. Phys.* **16**, 9847–9862 (2016).
55. Chen, H. et al. Understanding the joint impacts of fine particulate matter concentration and composition on the incidence and mortality of cardiovascular disease: A component-adjusted approach. *Environ. Sci. Technol.* **54**, 4388–4399 (2020).
56. Ou, Y., West, J. J., Smith, S. J., Nolte, C. G. & Loughlin, D. H. Air pollution control strategies directly limiting national health damages in the US. *Nat. Commun.* **11**, 957 (2020).
57. Dedoussi, I. C., Eastham, S. D., Monier, E. & Barrett, S. R. H. Premature mortality related to United States cross-state air pollution. *Nature* **578**, 261–265 (2020).
58. Li, C. et al. Trends in chemical composition of global and regional population-weighted fine particulate matter estimated for 25 years. *Environ. Sci. Technol.* **51**, 11185–11195 (2017).
59. Chan, E. A. W., Gantt, B. & McDow, S. The reduction of summer sulfate and switch from summertime to wintertime PM_{2.5} concentration maxima in the United States. *Atmos. Environ.* **175**, 25–32 (2018).
60. Pebesma, E. J. & Bivand, R. S. *Classes and methods for spatial data in R*. *R News*. Vol. 5, p. 9–13 (2005).
61. Original S code by Richard A. Becker and Allan R. Wilks. *R version by Ray Brownrigg. Enhancements by Thomas P Minka and Alex Deckmyn. maps: Draw Geographical Maps* <https://CRAN.R-project.org/package=maps> (2018).
62. Bivand, R. & Lewin-Koh, N. *maptools: Tools for Handling Spatial Objects* <https://CRAN.R-project.org/package=maptools> (2019).
63. Chamblee, J. *Geocodes USA with Counties* <https://data.healthcare.gov/dataset/Geocodes-USA-with-Counties/52wv-g36> (2013).
64. Spada, N. J. & Hyslop, N. P. Comparison of elemental and organic carbon measurements between IMPROVE and CSN before and after method transitions. *Atmos. Environ.* **178**, 173–180 (2018).
65. Centers for Disease Control and Prevention National Center for Health Statistics. *Compressed Mortality File 1999-2016 on CDC WONDER Online Database, released June 2017. Data are from the Compressed Mortality File 1999-2016 Series 20 No.2U, 2016, as compiled from data provided by the 7 vital statistics jurisdictions through the Vital Statistics Cooperative Program*. Accessed 22 March 2020, <http://wonder.cdc.gov/cmfi-icd10.html> (2017).
66. Centers for Disease Control and Prevention National Center for Health Statistics. *Compressed Mortality File: Age-Adjusted Rates*. <https://wonder.cdc.gov/wonder/help/cmfi.html#Age-Adjusted%20Rates> (2020).
67. Otte, T. L. & Pleim, J. E. The Meteorology-Chemistry Interface Processor (MCIP) for the CMAQ modeling system: updates through MCIPv3.4.1. *Geosci. Model Dev.* **3**, 243–256 (2010).
68. Remington, P. L., Catlin, B. B. & Gennuso, K. P. The County Health Rankings: rationale and methods. *Popul. Health Metr.* **13**, 11 (2015).
69. University of Wisconsin Population Health Institute (UWPHI). *County Health Rankings 2018 National Data* <https://www.countyhealthrankings.org/explore-health-rankings-rankings-data-documentation> (2018).
70. Wood, S. N. Thin plate regression splines. *J. R. Stat. Soc. B. Met.* **65**, 95–114 (2003).
71. Wood, S. N. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J. Am. Stat. Assoc.* **99**, 673–686 (2004).
72. R Core Team. *R: A language and environment for statistical computing* <https://www.R-project.org/> (2020).
73. U.S. Environmental Protection Agency. *Air Quality System (AQS)* <https://www.epa.gov/aqs> (2021).
74. Pye, H. O. T. *Data For Secondary Organic Aerosol And Cardiorespiratory Disease Mortality* <https://doi.org/10.5281/zenodo.5713903> (2021).

Acknowledgements

The U.S. EPA through its Office of Research and Development supported the research described here. It has been subjected to Agency administrative review and approved for publication but may not necessarily reflect official Agency policy. We thank Wayne Cascio, Chris Nolte, and Kristen Rappazzo for useful discussion. K.M.S. was supported by the Oak Ridge Institute for Science and Education (ORISE) Research Participation Program for the U.S. Environmental Protection Agency (EPA).

Author contributions

H.O.T.P. conceived the study. H.O.T.P. processed CMAQ output and wrote statistical model analysis code with input from C.K.W.C.; H.O.T.P. and B.N.M. developed and implemented CMAQ algorithms. K.W.A. performed CMAQ simulations. K.M.S. and K.W.A. processed and analyzed observed concentration data. H.O.T.P. prepared main figures and drafted paper. All authors provided input on the manuscript. All authors have given approval to the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-27484-1>.

Correspondence and requests for materials should be addressed to Havalva O. T. Pye.

Peer review information *Nature Communications* thanks Xavier Querol, Colleen Reid and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2021