Relationships Between Fine Particulate Air Pollution, Cardiometabolic Disorders, and Cardiovascular Mortality

C. Arden Pope III, Michelle C. Turner, Richard T. Burnett, Michael Jerrett, Susan M. Gapstur, W. Ryan Diver, Daniel Krewski, Robert D. Brook

Rationale: Growing evidence suggests that long-term exposure to fine particulate matter (PM$_{2.5}$) air pollution contributes to risk of cardiovascular disease (CVD) morbidity and mortality. There is uncertainty about who are most susceptible. Individuals with underlying cardiometabolic disorders, including hypertension, diabetes mellitus, and obesity, may be at greater risk. PM$_{2.5}$ pollution may also contribute to cardiometabolic disorders, augmenting CVD risk.

Objective: This analysis evaluates relationships between long-term PM$_{2.5}$ exposure and cardiometabolic disease on risk of death from CVD and cardiometabolic conditions.

Methods and Results: Data on 669,046 participants from the American Cancer Society Cancer Prevention Study II cohort were linked to modeled PM$_{2.5}$ concentrations at geocoded home addresses. Cox proportional hazards regression models were used to estimate adjusted hazards ratios for death from CVD and cardiometabolic diseases based on death-certificate information. Effect modification by pre-existing cardiometabolic risk factors on the PM$_{2.5}$-CVD mortality association was examined. PM$_{2.5}$ exposure was associated with CVD mortality, with the hazards ratios (95% confidence interval) per 10 μg/m$^2$ increase in PM$_{2.5}$ equal to 1.12 (1.10–1.15). Deaths linked to hypertension and diabetes mellitus (mentioned on death certificate as either primary or contributing cause of death) were also associated with PM$_{2.5}$. There was no consistent evidence of effect modification by cardiometabolic disease risk factors on the PM$_{2.5}$-CVD mortality association.

Conclusions: Pollution-induced CVD mortality risk is observed for those with and without existing cardiometabolic disorders. Long-term exposure may also contribute to the development or exacerbation of cardiometabolic disorders, increasing risk of CVD, and cardiometabolic disease mortality. (Circ Res. 2015;116:108-115. DOI: 10.1161/CIRCRESAHA.116.305060.)

Key Words: air pollution, epidemiology ■ metabolic syndrome X ■ particulate matter,
sex, and family history.19,20 The present analysis is based on the American Cancer Society (ACS) Cancer Prevention Study II (CPS-II) cohort, which collected data that included information related to multiple cardiometabolic risk factors by questionnaire at enrollment. The ACS CPS-II cohort data are linked with death certificate records and with modeled PM$_{2.5}$ exposure concentrations at residential address.

The objective of this analysis is to evaluate possible joint relationships between long-term PM$_{2.5}$ exposures and cardiometabolic disease on risk of death from CVD systematically. Specifically, we evaluate whether observed associations of long-term exposure to PM$_{2.5}$ air pollution on CVD mortality are greater in individuals with cardiometabolic risk factors, based on questionnaire information provided at time of cohort enrollment. Moreover, we explored the effects of long-term PM$_{2.5}$ exposure on mortality because of cardiometabolic disease (hypertension and diabetes mellitus) per se based on cause-of-death information available on death certificates.

Methods

An expanded Methods is available in the Online Data Supplement.

Study Population

The ACS CPS-II prospective cohort included 1184587 participants, enrolled by >77000 volunteers between September 1982 and February 1983. Participants were largely friends and family members of the volunteers and were recruited from all 50 states, the District of Columbia, and Puerto Rico. Participants in the ACS CPS-II cohort had to be aged ≥30 years and have ≥45 years. At enrollment, participants completed a 4-page, self-administered questionnaire providing their residential address and information on a range of demographic, lifestyle, medical, and other factors. The Emory University School of Medicine Human Investigations Committee provided ethics approval for the ACS CPS-II; ethics approval for the present analysis was obtained from the Ottawa Hospital Research Ethics Board. The ACS CPS-II cohort data are linked with death certificate records and with modeled PM$_{2.5}$ exposure concentrations at residential address.

The objective of this analysis is to evaluate possible joint relationships between long-term PM$_{2.5}$ exposures and cardiometabolic disease on risk of death from CVD systematically. Specifically, we evaluate whether observed associations of long-term exposure to PM$_{2.5}$ air pollution on CVD mortality are greater in individuals with cardiometabolic risk factors, based on questionnaire information provided at time of cohort enrollment. Moreover, we explored the effects of long-term PM$_{2.5}$ exposure on mortality because of cardiometabolic disease (hypertension and diabetes mellitus) per se based on cause-of-death information available on death certificates.

Exposure Estimates

Exposure to PM$_{2.5}$ was estimated by linking geocoded home addresses of the study participants to ambient PM$_{2.5}$ concentrations derived using a national-level hybrid land use regression (LUR) and Bayesian Maximum Entropy (BME) interpolation model (LUR-BME) described elsewhere.22 (Details on the geocoding of residential locations to be used in exposure assessment are presented in the Online Data Supplement.) Briefly, monthly PM$_{2.5}$ data for a total of 1464 monitoring sites from 1999 to 2008 were used to estimate the LUR-BME model in 2 stages. A training data set of 1329 monitors was used to select the variables, and 10% or 135 monitors were retained approximately for cross-validation. (Details on the model cross-validation analysis are presented in the Online Data Supplement.) In the first stage, the model was fit with a deterministic LUR with monthly pollution averages as the dependent variable and land use information as predictors. The LUR model used a deletion/substitution/addition algorithm with v-fold cross-validation to select predictor variables. This method reduces the chance of overfitting by continuously predicting on leave-out folds, meaning selected variables minimize the mean square error of predictions on data that are not used to fit the model. On the basis of this method, we selected 2 variables: traffic-weighted roads within 1000 m of a monitor (based on modeled traffic counts) and the cube of percentage of green space within a 100 m buffer around the monitor. In the second stage, a BME kriging interpolation model was used to capture the residual spatiotemporal variation in PM$_{2.5}$ concentrations not predicted in the first stage using LUR. Cross-validation resulted in an $R^2$ of 0.79. Scatter plots that illustrate the cross-validation prediction for the models, maps that illustrate the spatial variability of estimated exposures, and further documentation has been published elsewhere22 and are summarized in the Online Data Supplement. Monthly values from 1999 to 2004 were averaged and assigned to study participants using their geocoded addresses. The estimated overall mean PM$_{2.5}$ exposure concentration was 12.6 (SD=2.9) µg/m$^3$, with a range from 1 to 28 µg/m$^3$.

Statistical Analysis

Adjusted mortality hazard ratios (HRs) were estimated using the Cox proportional hazards regression models. Follow-up time in days was used as the time axis since enrollment. Survival times of those still alive at the end of follow-up were censored and, in analyses of cause-specific mortality, if death occurred for another cause, survival times were censored at the time of death. The models included the estimated PM$_{2.5}$ exposure concentration as a continuous variable. The models also controlled for multiple individual-level covariates as detailed elsewhere. Briefly, all models were stratified by 1-year age categories, sex, and race (white, black, and other), allowing each category to have its own baseline hazard. The individual-level covariates incorporated in the models, based on information collected from the 1982 ACS CPS-II enrollment questionnaire, included 13 variables that characterized current and former smoking habits (including smoking status of never, former, or current smoker, linear and squared terms for years smoked and cigarettes smoked per day, indicator for starting smoking at aged <18 years, and pipe/cigar smoker); 1 continuous variable that assessed exposure to second-hand cigarette smoke (hours/d exposed); 7 variables that reflected workplace PM$_{2.5}$ exposure in each subject’s main lifetime occupation; a variable that indicated self-reported exposure to dust and fumes in the workplace; variables that represented marital status (separated/divorced/widowed or single versus married); variables that characterized the level of education (high school, more than high school versus less than high school); 2 body mass index variables (linear and squared terms for body mass index); variables that characterized the consumption of alcohol (beer, missing beer, wine, missing wine, liquor, and missing); and variables that indicated quartile ranges of dietary fat index and quartile ranges of a dietary vegetable/fruit/fiber index.

To evaluate the sensitivity of the results to control for geographical, social, economic, and environmental settings (contextual conditions), information and 130000 individuals with missing individual-level data. The final analytic cohort used in this analysis included 669046 participants. A total of 237201 participants, ≈35% of the initial study cohort, died during the 22-year follow-up period.
some models also included ecological covariates obtained from the 1990 Census of Population Long-Form for the subjects’ residential zip code area. These ecological covariates are more completely documented elsewhere and included median household income; percentage of people with <125% of poverty-level income; percentage of unemployed individual aged 21-65 years; percentage of adults with <12th grade education; and percentage of the population who were black or Hispanic. These ecological covariates were included in the models using both zip code level data and zip code deviations from the county means.

Baseline HRs associated with an increment of 10 μg/m3 of PM2.5 were estimated for all-cause, CVD, hypertension, and diabetes mellitus mortality. Two approaches were used to evaluate the effect modification of cardiometabolic risk factors at time of enrollment on the PM2.5–CVD mortality association. First, adjusted HRs (and 95% confidence intervals) for CVD mortality were estimated in relation to 3 key categorical indicators of cardiometabolic risk (diabetes mellitus, doctor diagnosed high blood pressure, and heart disease at time of enrollment) and categorically high and low PM2.5 concentrations (>75th percentile and <25th percentile). To test for additive interactions between PM2.5 exposure and key cardiometabolic risk factors formally, the relative excess risk because of interaction, the attributable proportion because of interaction and the synergy index were calculated using the MOVER method for the analysis of 4x2 tables as documented elsewhere.

The second approach to evaluate the effect modification of cardiometabolic risk factors estimated adjusted HRs associated with increments in PM2.5 (using PM2.5 as a continuous variable) for cardiovascular mortality, while stratifying by all cardiometabolic risk factors that were available based on information from the ACS CPS-II enrollment questionnaire. These risk factors include body mass index; doctor diagnosed high blood pressure, heart disease, and diabetes mellitus; exercise levels; vegetable/fruit/fiber and fat intake; and the use of medications including aspirin, heart medications, and diuretics. Because the likelihood of any individual in the cohort having any of the key risk factors at enrollment depends partially on age and smoking status at enrollment, indicators of cardiometabolic risk were cross-stratified with 4 age-at-enrollment and smoking status strata (never smokers, age <60 years; never smokers, age ≥60 years; ever smokers, age <60 years; ever smokers, age ≥60 years). To test whether the associations differ for different follow-up times, we conducted the analysis, stratified across strata relating to cardiometabolic risk factors for 3 different follow-up periods: 0 to 7, 7 to 14, and 14 to 22 years.

Specific cause-of-death analyses were conducted using primary and contributing cause-of-death information provided on the death certificate focusing on hypertensive disease, diabetes mellitus, and interactions with other cardiovascular causes of death.

### Results

Estimated HRs (and 95% confidence intervals) associated with a 10 μg/m3 elevation in PM2.5 for all-cause, CVD, various subcategories of CVD, and diabetes mellitus mortality are presented in Table 1 and illustrated in Figure 1A. Estimates from these baseline Cox proportional hazards regression models indicated that all-cause and CVD mortality were significantly associated with long-term exposure to PM2.5. The largest and most statistically robust PM2.5–CVD mortality associations were with ischemic heart disease mortality, hypertensive disease and diabetes mortality, and mortality from a cause-of-death grouping that includes heart failure, cardiac arrest, and related complications.

Table 1. HRs (95% CI) Per 10-μg/m3 Increment in PM2.5 for All-Cause, Cardiovascular, and Diabetes Mellitus Mortality Using the Cox Model With Individual-Level Covariates, Without and With Ecological Covariates, and With Exposure Estimated Using the LUR-BME Model, Along With Number of Deaths and Relevant ICD-9 and ICD-10 Codes

<table>
<thead>
<tr>
<th>Primary Cause of Death</th>
<th>No. of Deaths</th>
<th>ICD-9 Codes</th>
<th>ICD-10 Codes</th>
<th>HRs (95% CIs) Per 10-μg/m3 PM2.5</th>
<th>Cox Model With Individual-Level Covariates; Exposure Based on LUR-BME</th>
<th>Cox Model With Individual-Level Plus Ecological Covariates; Exposure Based on LUR-BME</th>
</tr>
</thead>
<tbody>
<tr>
<td>All-cause mortality</td>
<td>237 201</td>
<td>100 149</td>
<td>45 644</td>
<td>453 1314</td>
<td>1.07 (1.05–1.09)</td>
<td>1.07 (1.06–1.09)</td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>100 149</td>
<td>390–459</td>
<td>410–414</td>
<td>420–429</td>
<td>1.14 (1.12–1.17)</td>
<td>1.12 (1.10–1.15)</td>
</tr>
<tr>
<td>Ischemic heart disease</td>
<td>45 644</td>
<td>410–414</td>
<td>420–429</td>
<td>430–438</td>
<td>1.19 (1.15–1.23)</td>
<td>1.14 (1.10–1.18)</td>
</tr>
<tr>
<td>Heart failure, cardiac arrest, and related</td>
<td>183 1314</td>
<td>420–429</td>
<td>420–429</td>
<td>420–429</td>
<td>1.12 (1.07–1.18)</td>
<td>1.11 (1.05–1.18)</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td>17 065</td>
<td>430–438</td>
<td>450–459</td>
<td>401–405</td>
<td>1.04 (0.99–1.10)</td>
<td>1.11 (1.05–1.17)</td>
</tr>
<tr>
<td>Hypertensive disease</td>
<td>312 90</td>
<td>401–405</td>
<td>401–405</td>
<td>401–405</td>
<td>1.20 (1.06–1.35)</td>
<td>1.11 (0.97–1.26)</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>4890</td>
<td>250</td>
<td>401–405</td>
<td>401–405</td>
<td>1.08 (0.97–1.19)</td>
<td>1.13 (1.02–1.26)</td>
</tr>
</tbody>
</table>

CI indicates confidence interval; HR, hazard ratio; ICD, International Classification of Diseases; LUR-BME, land use regression-Bayesian Maximum Entropy; and PM2.5, fine PM <2.5 μm in aerodynamic diameter.
not substantively different when the CVD mortality HRs (95% confidence interval) per 10 μg/m³ PM₂.₅ across strata relating to cardiometabolic risk factors were estimated after excluding those who reported taking blood pressure, heart, or diuretic medication or when estimated for different follow-up times.

Table 2 and Figure 1B and 1C present the HRs (95% confidence intervals) per 10 μg/m³ increase in PM₂.₅ for deaths when the death certificate indicated hypertensive disease or diabetes mellitus or both as primary or contributing causes of death. Deaths with either hypertension or diabetes mellitus as a primary or contributing cause were significantly associated with long-term PM₂.₅ exposure. Deaths with any indication of hypertension or diabetes mellitus on the death certificate were also stratified by primary cause of death. In general, the associations of PM₂.₅ with risk of death when hypertension or diabetes mellitus were mentioned on the death certificate (as either primary or contributing causes of death) in combination with each other or with other CVD causes of mortality are relatively large and statistically significant (Figure 1B and 1C). Although these results are constrained by the number of deaths (Table 2), they are suggestive of effects of long-term exposure to PM₂.₅ for CVD deaths that are linked to hypertension and diabetes mellitus.

**Discussion**

The present study is based on 22 years of follow-up of the ACS CPS-II prospective cohort, coupled with enhanced exposure assessment that provided estimates of ambient residential PM₂.₅ concentrations using linkage to the geocoded home addresses. As such, this analysis encompasses a large cohort, with a long follow-up time, with a large number of deaths, and with improved exposure spatial acuity. The results of this analysis corroborate previous findings of statistically robust associations of PM₂.₅ with all-cause and CVD mortality observed in the ACS CPS-II cohort, a finding also consistent with those from other cohorts.

The present study evaluated potential joint relationships between long-term PM₂.₅ exposures and cardiometabolic disease on risk of CVD mortality. Although showing that subjects with both high PM₂.₅ exposure and a pre-existing condition are at the highest risk of CVD mortality, the increased risk of CVD death associated with PM₂.₅ is similar in subjects with a
pre-existing condition and those without. This lack of effect modification may be influenced by inevitable imperfect classification of pre-existing cardiometabolic disorders based on questionnaire data collected at enrollment when participants were younger and at much lower risk of death. However, the absence of effect modification was observed both in participants aged ≥60 years, as well as those who were aged <60 years. In addition, there may be a concern about collider-stratification bias.26,27 For example, if PM$_{2.5}$ contributes to CVD mortality by increasing blood pressure, stratifying by hypertension removes this pathogenic pathway, and the PM$_{2.5}$–CVD mortality estimate stratified by hypertension reflects the effect of PM$_{2.5}$, independent of hypertension. Nevertheless, our results provide evidence of a PM$_{2.5}$–CVD mortality association across nearly all strata evaluated and do not substantiate previous reports, suggesting that underlying cardiometabolic diseases predispose individuals to the adverse health effects of PM$_{2.5}$, at least with respect to fatal events induced by chronic exposures.4,11

Alternatively, our findings suggest a less-considered mechanism where PM$_{2.5}$ influences the development of cardiometabolic disorders. Long-term PM$_{2.5}$ exposure was associated with a significant 34% increase in deaths linked to hypertension (any mention on death certificate), and with a 20% increase in deaths with hypertension as the primary cause. Previous studies have shown that short-term exposure to PM$_{2.5}$ can cause a rapid elevation in blood pressure.4,28 Indeed, emergency department visits specifically for hypertension have been shown to increase in relation to recent air pollution levels.29 There is also accruing evidence from epidemiological studies internationally that higher particulate pollution levels are associated with an increased incidence of hypertension.28,30 We recently demonstrated that long-term PM$_{2.5}$ exposure is associated with chronic hypertension (13% elevation per 10 μg/m$^3$ among 35 303 adults living in Ontario, Canada).31 These associations have some biological plausibility in that several pathways have been shown to be involved in both human and animal studies, including autonomic imbalance, systemic inflammation, and endothelial dysfunction.28

The results of this study are also consistent with recent evidence that PM$_{2.5}$ is associated with diabetes mellitus mortality. A 10-μg/m$^3$ increase in PM$_{2.5}$ was associated with an 18% increase in diabetes mellitus–related mortality. A growing body of evidence also supports the notion that not only can exposure acutely perturb glycemic control but also living in regions with higher PM$_{2.5}$ levels is associated with increased diabetes mellitus.12,30 For example, the incidence of diabetes mellitus increased by 8% to 11% per 10 μg/m$^3$ among 62 012 adults living in Ontario, Canada.14 The present results corroborate previous data from a national cohort demonstrating a more robust increase (49% increase in risk per 10 μg/m$^3$) in diabetes mortality.
mellitus–related mortality. Several human studies and animal experiments have elucidated viable biological mechanisms, whereby PM$_{2.5}$ can be capable of impairing metabolic insulin sensitivity via autonomic and systemic proinflammatory and oxidative stress pathways.

Our findings cannot explain precisely the mechanisms behind the association of PM$_{2.5}$ with increased mortality related to hypertension and diabetes mellitus. PM$_{2.5}$ may be triggering CVD deaths that are unrelated to hypertension or diabetes mellitus among individuals with previously developed hypertension and diabetes mellitus. This is suggested by the significant increase in several CVD-related causes of death listed as the primary cause among those with either hypertension or diabetes mellitus appearing anywhere on the death certificate. It is also possible that PM$_{2.5}$ might exacerbate the underlying hypertensive or diabetic disease state in a subacute manner during a period of weeks to months, thereby promoting a fatal event principally related to these disorders (hypertensive or hyperglycemic emergencies). Underlying this acute event, long-term PM$_{2.5}$ exposure might compromise chronic diabetic and blood pressure control (or contribute to their actual onset), and thereby promote the future fatal event related to these chronic disorders per se. It is likely that combinations of these 3 pathways are occurring at the population level.

This analysis has notable strengths. It uses data from a large, nation-wide, well-characterized, and managed prospective cohort followed up for 22 years. It uses a state-of-the-art LUR-BME model to provide exposure estimates for a relatively large number of cohort members, and it provides estimates of exposure with substantially improved spatial acuity. As documented elsewhere, cross-validation analysis demonstrated strong agreement between estimated exposures using the LUR-BME model and observed data ($R^2$=0.79) with no indications of bias or exceptional outliers.

This study also has limitations. Because the ACS CPS-II cohort included many friends and family members of ACS volunteers, it was not a random sample, but over-represented relatively affluent, well-educated individuals. As discussed elsewhere as part of a review and comparison of results from various cohort studies, there is some evidence that air pollution has a bigger effect on the less affluent and less educated. Therefore, pooled estimates from the ACS CPS-II cohort may underestimate the effect relative to the general population. Estimates of PM$_{2.5}$ exposure are not available throughout the lives of cohort members, or even for the full cohort follow-up period. Individual-level risk-factor data, including information relating to pre-existing cardiometabolic risk factors, and residential address, are based on information from the ACS CPS-II enrollment questionnaires, with no subsequent follow-up. The lack of air pollution and risk factor follow-up data reduces the precision of control for risk factors, constrains the ability to differentiate time dependency, and introduces some exposure misclassification that may bias the results toward the null.

Another key limitation of this study pertains to the use of cause-of-death information provided on death certificates. Death certificate information is inevitably incomplete on...
causes of death and comorbidities contributing to death and is clearly limited with respect to defining deaths caused by cardiometabolic disease states. With regards to cardiometabolic disease, mention of diabetes mellitus or hypertension on the death certificate is suggestive but far from definitive. Furthermore, many CVD deaths would likely include a significant number of individuals with cardiometabolic disease but did not have any mention of diabetes mellitus or hypertension on the death certificate. Similarly, because health information was collected only on the cohort enrollment questionnaire and on death certificates, subjects who developed cardiometabolic disorders during follow-up, but are still living, are not captured as events in this analysis at all. More definitive evaluations of joint contributions of air pollution and cardiometabolic disease with CVD mortality would require follow-up that also includes prospective tracking of indicators of the development and progression of cardiometabolic disease.

In conclusion, cardiometabolic disorders are common and, because of increasing obesity, sedentary lifestyles, and atherogenic diets, are growing in prevalence worldwide.3,9,20,33 The prominent role of CVD, hypertension, diabetes mellitus, and air pollution in contributing to the global burden of disease has been well documented.3,35 The potential global public health implications of joint relationships between fine PM air pollution, cardiometabolic disorders, and cardiovascular mortality are substantial.

Sources of Funding
The American Cancer Society funded the creation, maintenance, and follow-up of the Cancer Prevention Study II cohort. C. Arden Pope III was funded, in part, by the Mary Lou Fulton Professorship, Brigham Young University. Michelle C. Turner was funded by a Government of Canada Banting Postdoctoral Fellowship. Daniel Krewski is the Natural Sciences and Engineering Research Council (NSERC) Chair in Risk Science at the University of Ottawa. Michael Jerrett was supported for the exposure assessment modeling funding from NIEHS (National Institute of Environmental Health Sciences) R01 ES019573-01, the Centers for Disease Control and Environmental Public Health Tracking and Health Canada.

Disclosures
None.

References


29. Szyszkowicz M, Rowe BH, Brook RD. Even low levels of ambient air pollutants are associated with increased emergency department visits for hypertension. *Can J Cardiol*. 2012;28:360–366.


---

**Novelty and Significance**

**What Is Known?**

- Long-term exposure to fine particulate air pollution is associated with risk of cardiovascular mortality.
- It is unknown who is most susceptible but those with underlying cardiometabolic disorders may be at greater risk.

**What New Information Does This Article Contribute?**

- Cardiovascular mortality–pollution associations were observed for those with and without existing cardiometabolic disorders ascertained at study baseline.
- Deaths with hypertension or diabetes mellitus mentioned on death certificate were more strongly associated with air pollution than deaths without these conditions mentioned, suggesting that long-term exposure may contribute to cardiometabolic disorders augmenting cardiovascular disease risk.

---

This analysis used data from a large, nation-wide cohort, with >20 years of follow-up, with enhanced exposure assessment, and with control for multiple key individual and geospatial contextual variables. Associations between mortality risk and elevated exposures to fine particulate air pollution were remarkably robust. Pre-existing cardiometabolic risk factors (based on cohort enrollment information) did not modify the effect of pollution exposure on cardiovascular mortality. Deaths linked to hypertension or diabetes mellitus (based on information from death certificates), however, were more strongly associated with pollution, suggesting that long-term pollution exposure may contribute to the development or exacerbation of cardiometabolic disorders, increasing risk of cardiovascular and cardiometabolic disease mortality.
Relationships Between Fine Particulate Air Pollution, Cardiometabolic Disorders, and Cardiovascular Mortality

C. Arden Pope III, Michelle C. Turner, Richard T. Burnett, Michael Jerrett, Susan M. Gapstur, W. Ryan Diver, Daniel Krewski and Robert D. Brook

Circ Res. 2015;116:108-115; originally published online October 27, 2014;
doi: 10.1161/CIRCRESAHA.116.305060

Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/116/1/108

Data Supplement (unedited) at:
http://circres.ahajournals.org/content/suppl/2014/10/27/CIRCRESAHA.116.305060.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation Research can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation Research is online at:
http://circres.ahajournals.org/subscriptions/
SUPPLEMENTAL MATERIAL

EXPANDED METHODS

Study Population

This analysis is based on data from the American Cancer Society (ACS) Cancer Prevention Study II (CPS-II) cohort. This cohort has been used previously to study the health effects of air pollution.\textsuperscript{1,2} The ACS CPS-II prospective cohort included 1,184,587 participants who were enrolled by over 77,000 volunteers between September 1982 and February 1983. Participants were largely friends and family members of the volunteers, and were recruited from all 50 states, the District of Columbia, and Puerto Rico. Participants in the ACS CPS-II cohort were at least 30 years of age and had at least one family member aged 45 years or older. The Emory University School of Medicine Human Investigations Committee provided ethics approval for the ACS CPS-II; ethics approval for the present analysis was obtained from the Ottawa Hospital Research Ethics Board.

At time of enrollment, participants completed a four-page, self-administered questionnaire providing their residential address and information on birth date, age, height and weight, and marital status. Additionally the questionnaire collected information on family history in relation to cancer, history of diseases, current physical condition, smoking history and habits, diet (including consumption of beer, wine, and hard liquor), use of medications and vitamins, occupational history and exposures, and additional miscellaneous information. Copies of the full ACS CPS-II questionnaires for both men and women are available at (http://www.cancer.org/research/researchprograms/funding/epidemiology-cancerpreventionstudies/studyquestionnaires/index).

Vital status follow-up was conducted every two years. For the years 1984, 1986, and 1988, vital status was obtained by the study volunteers and confirmed by obtaining death certificates. Subsequent vital status follow-ups were conducted using computerized record linkage to the National Death Index.\textsuperscript{3} In this analysis participants were followed up from the time of enrollment (between September 1982 and February 1983) through 2004. During the first six years of follow-up, cause of death was coded using a 2-digit ACS CPS-II code that was a consolidation of International Classification of Diseases, Ninth Revision (ICD-9) codes. Subsequent cause of death coding used ICD-9 or ICD-10 codes. Cause of death was captured for the underlying or primary cause of death, and the next two contributing causes of death, unless a cancer was reported later in the death certificate. In the event of a cancer reported later on a death certificate it would be captured instead of the second contributing cause of death.

Although the total ACS CPS-II cohort included nearly 1.2 million participants, approximately 385,000 individuals with invalid home address information and 130,000 individuals with missing individual-level data were excluded. The final analytic cohort used in this analysis included 669,046 participants. A total of 237,201 participants, approximately 35% of the initial study cohort, died during the 22-year follow-up period.

Geocoding of Residential Locations

In preparation for geocoding of the ACS-CPS-II cohort, Dr. Michael Jerrett, UC Berkeley, accompanied Mr. Zev Ross, ZevRoss Spatial Analysis, to Atlanta and met with ACS professional staff to ensure smooth access to the data and correct geocoding procedures. Geocoding was performed on the ACS CPS-II cohort data with a Dell Precision M65 laptop at the ACS offices in Atlanta. The laptop included ArcGIS 9.3.1 and ZP4 (Semaphore Corporation) software for use in the geocoding (expiration date of July 2010). The road data and geocoding locators were prepared in advance and TeleAtlas-based
StreetMap data from ESRI were used for all geocoding. A single composite locator made up of two, very similar street locators was created. Both individual locators used a side offset of 15 meters. The default end offset of 3 m, spelling sensitivity of 80%, a minimum candidate score of 10 and a minimum match score of 60 were used and was set to assign a non-match to addresses that matched more than one address. The only difference between the two locators used was that the primary locator (“Street_Address”) used 2005 TeleAtlas roads (this is the ESRI 9.3.1 Update to the Data & Maps) while the other (“Street_Address1”) used 2003 TeleAtlas roads. Each one of these road files includes about 40 million road segments. In our testing, we noticed that a small percentage of addresses (<1%) were properly geocoded with the older road network only and decided that it would be valuable to include a locator based on this older street network to capture this small percentage of additional matches.

**Testing in advance of geocoding.** We tested for address correction and geocoding speed and accuracy using voter registration data from three states (NY, WA and PA). Voter registration data provides ideal testing data in the sense that addresses are likely to be a little messy and it represents one of the few publicly available datasets with millions of records of real residential addresses. The drawback of voter registration data is that only a limited number of states make the data available for free and, as a result, the accuracy and speed tests will not necessarily be representative of a national dataset. From a speed perspective, voter registration data was likely to be deceptively fast because it consists of geographically close addresses reducing the amount of search time needed, in particular, the address correction software. We tested the address processing on a sample of 3 million records from the states above. In addition, we conducted a qualitative comparison of geocoding accuracy using 100 manually selected addresses in New York City for which we could identify the actual building/address on an aerial photo that has been orthographically corrected. We chose to look at NYC because of the high number of verifiable addresses, high-resolution aerial photography available for the area and because Queens County has an unusual address numbering system (hyphenated) that we wanted to test with the address correction and geocoding methodology. Overall, we found that the locators based on TeleAtlas-based StreetMap data performed extremely well on voter registration data. Match rates were high, matches appeared to be highly accurate and the speed was impressive. Both address correction and geocoding could be completed less than 2 hours each.

**Address raw data.** ACS professional staff provided tab-delimited text files for geocoding. These include a file for all CPS-II baseline cohort participants (all82.txt) with 1,175,991 records. The file included the variables ID1, street, city, state and zip code. The address history dataset also includes a field for the day, month and year of date archive.

**Manual** address processing. The files were read into R statistical software (version 2.9.0) and reviewed using regular expressions scripting to identify all characters that were not numbers or letters. Based on this review, several characters were identified for removal in advance of address correction (these include *, %, “, \, /, ), (^, `), $, /). A unique sequential ID (called zID2) was added because the ID1 is not unique in the address history file. Finally, all the addresses with “C/O” at the beginning were identified. These likely represent “Care Of” and would not be a home address. The “C/O” was stripped from the address and a TRUE/FALSE field to identify these later if necessary was appended.

**Address correction.** ZP4 software from Semaphore Corporation was used to conduct address correction. We used the April 2010 release (expiration date of July 2010) and included the add-on route-to-street conversion data (called “LACS” by Semaphore) to convert rural-style addresses to street-style addresses for improved geocoding accuracy. The program was run in batch mode using the GUI and a CSV input file. The address correction took more than 10 hours and was allowed to run overnight. The resulting address corrected file was imported into R and stripped of non-essential fields for geocoding.
Geocoding performance. As mentioned above, the geocoding was performed using a composite locator based on TeleAtlas street data available in ArcGIS Data & Maps. The geocoding was executed using a Python script run from the Command Prompt. Geocoding the ACS data took somewhat longer than our test data (approximately 2.5 hrs). To get a sense of which addresses should have been geocodeable, the geocodeable addresses were identified using the criteria below. In all cases, these are case insensitive. This is an approximate estimate of an address’s ability to be geocoded, some of these might actually be geocodeable and other addresses might not be geocodeable. A non-geocodeable address was defined as an address with one or more of the following:

1. Street address includes the word “box”
2. No letters in address
3. No numbers in address
4. An address beginning with “RR” followed by whitespace followed by a number.
5. An address with “RT” followed by whitespace followed by a number.
6. An address with no zip AND no city
7. An address with no zip AND no state

None of these were excluded from the geocoding but were identified to assess match rates. In total 84% for all 82 addresses were geocodeable. Among geocodeable addresses at least 89% of addresses were geocoded at a score of 80 or above. Positional match rates are assigned a score from 0 to 100. The score is generated from various address elements. Scores will be lower if the candidate address contains misspellings, incorrect information, addresses outside the range on a street segment, or missing elements of the address. Detailed information on match rates is presented elsewhere.5

Exposure estimates

Exposure model estimations. Exposure to PM$_{2.5}$ was estimated by using the geocoded home addresses of the study participants and linking them to ambient PM$_{2.5}$ concentrations derived using a national-level hybrid land use regression (LUR) and Bayesian Maximum Entropy (BME) interpolation model (LUR-BME) more fully documented elsewhere.6 Monthly PM$_{2.5}$ data for a total of 1,464 monitoring sites from 1999 to 2008 were used to estimate the LUR-BME model in two stages. A training data set of 1329 monitors was used to select the variables, and approximately 10% or 135 monitors were retained for cross-validation. In the first stage, the model was fit with a deterministic LUR with monthly pollution averages as the dependent variable and land use information as predictors. The LUR model used a deletion/substitution/addition algorithm with v-fold cross-validation to select predictor variables. This method reduces the chance of over fitting by continuously predicting on leave-out folds, meaning selected variables minimize the mean square error of predictions on data that are not used to fit the model. Based on this method two variables were selected: traffic-weighted roads within 1000m of a monitor (based on modeled traffic counts) and the cube of percentage of green space within a 100m buffer around the monitor. In the second stage, a BME kriging interpolation model was used to capture the residual spatiotemporal variation in PM$_{2.5}$ concentrations not predicted in the first stage using LUR. Monthly values from 1999 to 2004 were averaged and assigned to study participants using their geocoded addresses. The estimated overall mean PM$_{2.5}$ exposure concentration was 12.6 (SD=2.9) µg/m$^3$, with a range from 1 to 28 µg/m$^3$.

Cross-validation. The randomly selected cross-validation data of 135 monitors that were not used in the variable selection or model prediction were used to evaluate how well the model predicted at locations not used to calibrate the model. Supplemental Figure 1 compares cross-validation predictions of the LUR-BME plotted against the observed monthly data. This analysis showed good agreement between observed data and LUR-BME predictions with no significant bias or outliers. The model R$^2$ is 0.79, which suggests the model predicts well at locations different from those used to calibrate the model.
Supplemental Figure 2 shows a map of the LUR-BME models averaged over the entire study period. The patterns present have spatial patterns consistent with well-known regional patterns of PM$_{2.5}$ pollution, with areas of Central and Southern California having the highest levels, followed by parts of the Industrial Midwest and the Southeast. Supplemental Figure 3 shows a zoom in of the Los Angeles Metropolitan Area. Here areas with large highways are clearly visible indicating the ability of the model to predict small-area variation in PM$_{2.5}$.

**Statistical Analysis**

Adjusted mortality hazard ratios (HRs) were estimated using the Cox proportional hazards regression models. Follow-up time in days since enrollment was used as the time axis. Survival times of those still alive at end of follow-up were censored and, in analyses of cause-specific mortality, if death occurred for another cause, survival times were censored at the time of death. All models were stratified by one-year age categories, sex, and race (white, black, other), allowing each category to have its own baseline hazard. The age, sex, and race information was taken directly from the ACS CPS-II enrollment questionnaire. The models included the estimated PM$_{2.5}$ exposure concentration as a continuous variable, based on the LUR-BME exposure model described above.

The models also controlled for multiple individual-level covariates as follows. Tobacco smoke covariates included 13 variables that characterized current and former smoking habits (including smoking status of never, former, or current smoker, linear and squared terms for years smoked and cigarettes smoked per day, indicator for starting smoking at younger than 18 years of age, and pipe/cigar smoker) and one continuous variable that assessed exposure to second hand cigarette smoke (hours/day exposed). All of the tobacco smoke related covariates were obtained directly from self-reported information collected from the ACS CPS-II enrollment questionnaire. Covariates that controlled for occupational exposures included a variable that indicated regular occupational exposure to dust and fumes (including asbestos, chemical/acid/solvents, coal or stone dusts, coal tar/pitch/asphalt, diesel engine exhaust, or formaldehyde) as self-reported on the ACS CPS-II enrollment questionnaire and seven variables that reflected workplace exposure in each subject’s main lifetime occupation. The seven workplace exposure variables were indicator variables that indicated rankings of occupational exposures derived from occupational information from the ACS CPS-II enrollment questionnaire and documented elsewhere. Variables that represented marital status (separated/divorced/widowed or single versus married), variables that characterized levels of education (high school, more than high school versus less than high school), two body mass index (BMI) variables (linear and squared terms for BMI), and variables characterizing consumption of alcohol (beer, missing beer, wine, missing wine, liquor, missing liquor) were also included as covariates in the models and were obtained directly from self-reported information collected from the ACS CPS-II enrollment questionnaire. Dietary covariates included indicator variables for quartile ranges of a dietary fat index and quartile ranges of a dietary vegetable/fruit/fiber index. These diet indices were derived based on diet information provided in the ACS CPS-II enrollment questionnaire as documented elsewhere.

To evaluate the sensitivity of the results to control for geographical, social, economic, and environmental settings (contextual conditions), some of the Cox Proportional Hazards models also included ecologic covariates obtained from the 1990 Census of Population Long-Form for the subjects’ residential zip code area. These contextual variables included: median household income; percentage of people with < 125% of poverty-level income; percentage of unemployed persons over the age of 16 years; percentage of adults with less than 12th grade education; and percentage of the population who were Black or Hispanic. These ecological covariates were included in the models using both zip-code level data as well as zip-code deviations from the county means.
Baseline HRs associated with an increment of 10 µg/m³ of PM$_{2.5}$ were estimated for all-cause, CVD, hypertension and diabetes mortality. Two approaches were used to evaluate effect modification of cardiometabolic risk factors at time of enrollment on the PM$_{2.5}$-CVD mortality association. First, adjusted HRs (and 95% CIs) for CVD mortality were estimated in relation to three key categorical indicators of cardiometabolic risk (diabetes, doctor diagnosed high blood pressure, and heart disease at time of enrollment) and categorically high and low PM$_{2.5}$ concentrations (> 75th percentile and < 25 percentile). To formally test for additive interactions between PM$_{2.5}$ exposure and key cardiometabolic risk factors, the relative excess risk due to interaction, the attributable proportion due to interaction, and the synergy index were calculated using the MOVER method for the analysis of 4 x 2 tables as documented elsewhere.\textsuperscript{11}

The second approach to evaluate effect modification of cardiometabolic risk factor estimated adjusted HRs associated with increases in PM$_{2.5}$ (using PM$_{2.5}$ as a continuous variable) for cardiovascular mortality, while stratifying by all cardiometabolic risk factors that were available based on information from the ACS CPS-II enrollment questionnaire. These risk factors include: BMI levels, doctor diagnosed high blood pressure, heart disease, and diabetes, exercise levels; vegetable/fruit/fiber and fat intake; and use of medications including aspirin, heart medications, and diuretics. Specifically, different ranges of BMI (<25, 25-30, 30-35, 35+) were determined using height and weight information provided on the enrollment questionnaire. Doctor diagnosed blood pressure, heart disease, and diabetes along with exercise levels and medication use were all self-reported on the enrollment questionnaire. Quartile ranges of a dietary fat index and quartile ranges of a dietary vegetable/fruit/fiber index were derived based on diet information provided in enrollment questionnaire as noted above and more fully documented elsewhere (Chao et al. 2000). Because the likelihood of any individual in the cohort having any of the key risk factors at enrollment depends partially on age and smoking status at enrollment, indicators of cardiometabolic risk were cross-stratified with four age-at-enrollment and smoking status strata (never smokers, age < 60; never smokers, age >= 60; ever smokers, age < 60; ever smokers, age >= 60). In order to evaluate if the associations differ for different follow-up times, we conducted the analysis, stratified across strata relating to cardiometabolic risk factors for three different follow-up periods: 0-7, 7-14, and 14-22 years.

Specific cause-of-death analyses were conducted using primary and contributing cause-of-death information provided on the death certificate focusing on hypertensive disease, diabetes and interactions with other cardiovascular causes of death.
Supplemental References


Supplemental Figure I. Observed on Predicted Plot from PM$_{2.5}$ levels based on the LUR-BME Model; Data from 135 randomly selected cross-validation sites with monthly averages. Adapted from Beckerman et al. 2012.\textsuperscript{6}
Supplemental Figure II: Coterminal United States with Predicted PM$_{2.5}$ levels based on the LUR-BME Model. Adapted from Beckerman et al. 2012.\textsuperscript{6}
Supplemental Figure III: Los Angeles Metropolitan Area with Predicted PM$_{2.5}$ levels based on the LUR-BME Model. Adapted from Beckerman et al. 2012.$^6$