Estimating the Causal Effect of Low Levels of Fine Particulate Matter on Hospitalization

Maggie Makar,[a] Joseph Antonelli,[b] Qian Di,[c] David Cutler,[d] Joel Schwartz,[e] and Francesca Dominici[f]

Background: In 2012, the EPA enacted more stringent National Ambient Air Quality Standards (NAAQS) for fine particulate matter ($PM_{2.5}$). Few studies have characterized the health effects of air pollution levels lower than the most recent NAAQS for long-term exposure to $PM_{2.5}$ (now 12 μg/m^3).

Methods: We constructed a cohort of 32,119 Medicare beneficiaries residing in 5138 US ZIP codes who were interviewed as part of the Medicare Current Beneficiary Survey (MCBS) between 2002 and 2010 and had 1 year of follow-up. We considered four outcomes: all-cause hospitalizations, hospitalizations for circulatory diseases and respiratory diseases, and death.

Results: We found that increasing exposure to $PM_{2.5}$ from levels lower than 12 μg/m^3 to levels higher than 12 μg/m^3 is associated with increases in all-cause admission rates of 7% (95% CI = 3%, 10%) and in circulatory admission hazard rates of 6% (95% CI = 2%, 9%). When we restricted analysis to enrollees with exposure always lower than 12 μg/m^3, we found that increasing exposure from levels lower than 8 μg/m^3 to levels higher than 8 μg/m^3 increased all-cause admission hazard rates by 15% (95% CI = 8%, 23%), circulatory by 18% (95% CI = 10%, 27%), and respiratory by 21% (95% CI = 9%, 34%).

Conclusions: In a nationally representative sample of Medicare enrollees, changes in exposure to $PM_{2.5}$, even at levels consistently below standards, are associated with increases in hospital admissions for all causes and cardiovascular and respiratory diseases. The robustness of our results to inclusion of many additional individual level potential confounders adds validity to studies of air pollution that rely entirely on administrative data.

(*Epidemiology* 2017;28: 627–634)

To protect public health and welfare against the dangers of air pollution, the US Environmental Protection Agency (EPA) establishes National Ambient Air Quality Standards (NAAQS). In response to mounting evidence demonstrating the harmful effects of exposure to fine particulate matter, in 2012, the EPA enacted more stringent NAAQS for fine particulate matter ($PM_{2.5}$). As air pollution standards decrease, regulatory actions are becoming increasingly expensive with the annual cost of implementation and compliance with the NAAQS reaching tens of billions of dollars.[1,2] Although there are massive benefits to reduced air pollution levels[3,4] that far exceed their costs, research examining the public health benefits of cleaner air will be subjected to immense scrutiny due to the potential costs associated with more stringent regulatory policy. Despite a substantial amount of epidemiologic literature on the health effects of long-term exposure to air pollution,[5–13] few studies have characterized the health effects of air pollution at levels in accordance with or lower than the most recent NAAQS for long-term exposure to $PM_{2.5}$ (now set at 12 μg/m^3). From this point forward, when we refer to the NAAQS, we will be referring to the long-term standards for $PM_{2.5}$. Recent studies[14,15] have found positive associations between short-term exposure to air pollution and mortality, whereas another study[16] found a protective association of short-term $PM_{2.5}$ with exacerbation of chronic obstructive pulmonary disorder. Positive associations between long-term exposure to concentrations of

Submitted 10 January 2017; accepted 22 May 2017.

From the [a]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA; [b]Department of Biostatistics, Harvard T.H. Chan School of Public Health; [c]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA; [d]Department of Economics, Harvard University, Cambridge, MA; [e]Department of Environmental Health, Harvard T.H. Chan School of Public Health; and [f]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA.

Maggie Makar and Joseph Antonelli are co-lead authors.

Supported by the National Institutes of Health (Grants P01 CA134294, P50 MD010428, R01 GM111339, R01 ES024332, R01 ES026217, R21 ES024012, R35 CA197449); the Environmental Protection Agency (Grant 83587201-0); and the Health Effects Institute (Grant 4953-RFA14-3/16-4).

The authors report no conflicts of interest.

Sharing Health Outcomes: Medicare health outcome data will be stored on a highly secure server under the supervision of Dr. Yun Wang of the Harvard T.H. Chan School of Public Health. To allow for our analytical health outcome datasets to be replicated by researchers outside of our team, we will provide the following: (1) the list of Medicare files that we used; (2) SAS macros to efficiently process raw data files; and (3) simulated health outcome datasets that represent hypothetical patients and illustrate our data formatting conventions for our tools to be used by other research groups.

Sharing the Linked National Data using Dataverse: Instead of posting data on a private web server or developing ad-hoc data management solutions, all nonsensitive datasets (e.g., EPA AQS pollution data), simulated health outcomes datasets, replication instructions, and links to open-source software will be made publicly available.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Francesca Dominici, Department of Biostatistics, Harvard T.H. Chan School of Public Health, 655 Huntington Avenue, Building II, Room 441, Boston, MA 02115. E-mail: fdominic@hsph.harvard.edu.

ISSN: 1044-3983/17/2805-0627
DOI: 10.1097/EDE.0000000000000690

$PM_{2.5}$ mostly below 12 μg/m^3 and mortality were recently reported in a Canadian cohort.[17] Additionally, there has been little scientific literature examining the effects of air pollution in smaller cities, towns, rural areas and areas with sparse monitoring. As air pollution levels decrease, studies are needed to determine if further reductions will lead to substantial improvements in health.

In addition, traditional observational cohorts have modeled the outcome as a function of exposure and confounders. Provided that the confounder model is correctly specified (including no omitted confounders), such studies provide causal estimates of the effect of exposure, conditional on the covariates. More recent causal modeling approaches model exposure as a function of covariates, and conditional on the exposure model being correctly specified, can provide marginal estimates of the causal effects of exposure on outcome. Often this can be advantageous because many predictors of health (e.g., alcohol consumption) are not causes of air pollution, but are indirectly associated with it through a common cause, such as socioeconomic status. It may be easier to model the effect of income on exposure than the effect of alcohol on cardiovascular disease. We have applied one such causal modeling approach to our data.

In this study, we build upon the existing literature in several ways: (1) we use inverse probability weighting (IPW), enabling us to estimate: (a) the "causal" effects of increasing $PM_{2.5}$ levels from below 12 μg/m^3 to above 12 μg/m^3 and (b) the causal effects of increasing $PM_{2.5}$ from below 8 μg/m^3 to above 8 μg/m^3 but always below 12 μg/m^3; (2) we use estimates of fine particulate matter ($PM_{2.5}$) on a 1 × 1 km grid to compute exposure at the ZIP code level; (3) we use open cohort data from Medicare claims data, which allow us to enroll new individuals each year and examine the health effects over time as air pollution levels continue to decline; (4) we link Medicare claims data to data from the Medicare Current Beneficiary Survey (MCBS),[18] which provides information on an extensive list of individual level behavioral risk factors and allows us to control for important confounders such as body mass index (BMI) and smoking habits; and (5) we assess the sensitivity of our estimates of causal effects with respect to several modeling assumptions including the following: (a) restriction of our study population to individuals already exposed to low pollution levels (<12 μg/m^3) and most importantly (b) inclusion/exclusion of a large set of individual level behavioral risk factors (such as smoking and BMI) when we consider methods for confounding adjustment. Assessing the robustness of causal effects of air pollution to the lack of adjustment for these individual level behavioral risk factors is very important as these factors are generally hard to measure and only available from cohort studies.

METHODS

This study was approved by the Institutional Review Board from the Harvard T.H. Chan School of Public Health.

Medicare–Medicare Current Beneficiary Survey Cohort

We consider all Medicare fee-for-service enrollees who reside in the continental United States and participated in the Medicare Current Beneficiary Survey (MCBS) from 2002 to 2010. This allows us to construct an open cohort of N = 32,119 Medicare beneficiaries residing in 5138 unique ZIP codes. The MCBS is a representative survey of the Medicare population. It is designed as a rotating panel, where every MCBS participant is interviewed three times a year for a maximum of four consecutive years. For the purposes of this study, we only retained one interview per year, leading to a total of 68,789 unique patient-years. We defined the reference date to be the last interview date in a given year. Figure 2 shows the timeline and study design.

We excluded patients not enrolled in Medicare for the entire look back period and outcome observation period. Specifically, we excluded patients who are not yet enrolled in Medicare or ones who are enrolled in a Healthcare Maintenance Organization (HMO). We also excluded patients who reside in US outlying territories. Details regarding inclusion/exclusion criteria are described in Figure 1.

Low Pollution Cohort

We created a "low pollution cohort" that only includes those individuals from the full cohort whose exposure to $PM_{2.5}$ is lower than 12 μg/m^3 during the 2-year period prior to the reference date. This reduces the number of unique subjects included in the cohort from 32,119 to 18,144. The purpose of constructing the low pollution cohort is to assess if there is evidence of a causal effect of air pollution on health outcomes even among individuals with exposure levels that are already below the annual NAAQS. In particular, we will use this cohort to examine if there exists a further reduction in risk for subjects exposed to $PM_{2.5}$ less than 8 μg/m^3, which has been identified by previous work as a level with low risk.[19]

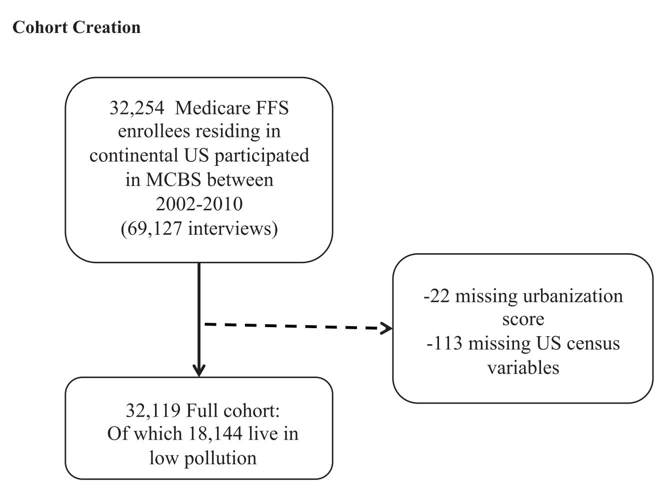

FIGURE 1. Inclusion criteria and cohort creation. Medicare FFS enrollees residing in continental US participating in the MCBS. FFS indicates fee-for-service.

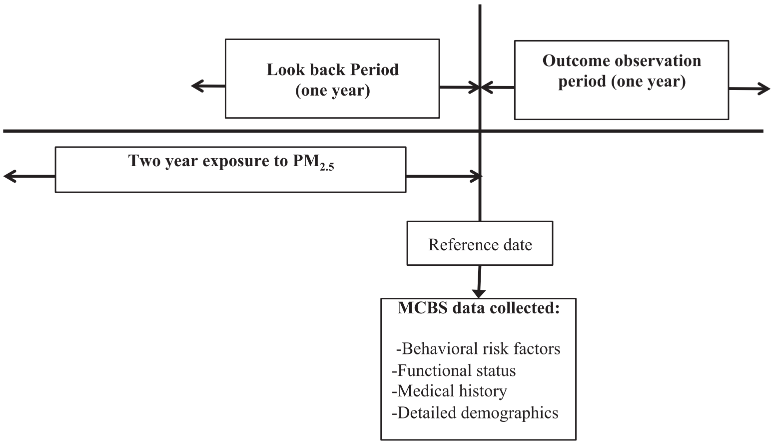

FIGURE 2. Data collection process for a hypothetical patient.

Study Design

Exposure to $PM_{2.5}$

To estimate daily levels of $PM_{2.5}$ for the entire study period (2002–2010) and for every ZIP code included in the study, we applied a previously developed exposure prediction model.[20] This model integrates satellite-based aerosol optical depth measurement, chemical transport model simulation, meteorologic variables, land-use terms, and other auxiliary variables. We trained this hybrid model to monitor $PM_{2.5}$ with a neural network. Neural networks account for nonlinearity and interactions between variables, thus improving model performance. We used the trained neural network to estimate daily $PM_{2.5}$ on a 1×1 km grid for the entire continental United States. We then estimated each individual's exposure to $PM_{2.5}$ by averaging $PM_{2.5}$ levels across space (from the 1×1 km grid to ZIP code of residence) and across time (for the 2 years prior to the reference date) (Figure 3). In previous work,[20] we reported a 10-fold cross validation of $R^2 = 0.84$ for daily measurements, at the monitoring sites, for the period 2000–2012, and for the entire continental United States. This indicates high correlation between predicted and monitored $PM_{2.5}$. This correlation is anticipated to be even higher when we aggregate these values across time (day to year) and across space (1×1 km grid cells to ZIP code). For further details of the exposure assessment, refer to Di et al.[20]

Outcome Observation Period

We identified a 1-year follow-up period from the reference date to ascertain health outcomes from the claims data (Medicare Provider Analysis and Review [MedPAR] part A). We considered the following: (1) all-cause mortality; (2) all-cause hospitalizations; (3) hospitalizations with a coded circulatory disease (International Classification of Diseases, Ninth Revision [ICD-9]: 390–459); and (4) hospitalizations with a coded respiratory disease (ICD-9: 460–519). Diagnoses, procedures, and outcomes are defined according to the highest level of the ICD-9 hierarchy.

Potential Confounders

Data extracted from multiple sources (listed below) provide information on a total of 122 potentially confounding factors. eTable 1 (http://links.lww.com/EDE/B215) summarizes the mean and standard deviation of all variables and outcomes in the study, separately for exposure higher and lower than 12 $\mu g/m^3$, respectively.

MCBS Data

For each enrollee in the MCBS–Medicare cohort, we extracted an extensive list of potential confounders from the MCBS data that are collected at the reference date. These include the following: patients' functional status (e.g., if they have difficulty walking), their behavioral risk factors (e.g., smoking status), and their detailed demographics (e.g., marital status and level of education) among others ($P = 73$), where with P we denote the total number of observed covariates.

Look Back Period

We extracted information from Medicare claims data on individual level comorbidity during the 1-year look back period. Specifically, from Medicare part A, we constructed several binary variables encoding the presence or absence of a number of procedures during hospitalization (e.g., operations on the digestive system) ($P = 27$). Basic patient demographics (e.g., age, race, gender, and mailing ZIP code) are collected from the Master Beneficiary Summary and the Denominator files ($P = 9$).

ZIP Code Level Data

Finally, we gather ZIP code level data including urbanization score as estimated by the US Department of Agriculture (USDA) ($P = 3$) and socioeconomic variables from the US Census ($P = 10$).[21]

Main Analysis

Throughout, we will be relying on three key assumptions necessary for making causal statements: the stable unit treatment value assumption, positivity, and the assumption of no unmeasured confounding. The stable unit treatment value assumption[22] assumes that the outcome of a given observational unit is not affected by the treatment assignment (i.e., exposure to high vs. low pollution levels) received by another unit. Positivity states that all experimental units have a positive

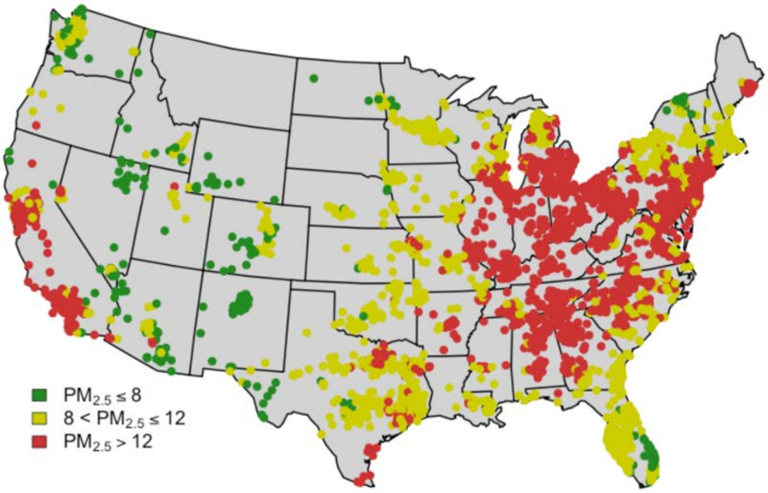

FIGURE 3. Average exposure in the year 2002 for each of the 5138 ZIP codes included in the study. These are estimated exposures as described in Di et al.[20]

probability of receiving each level of treatment (i.e., exposure to high or low levels of air pollution). We will assess this assumption by looking at propensity score overlap in eFigure 1 (http://links.lww.com/EDE/B215) and find that it is reasonable. Finally, no unmeasured confounding implies that our full set of available covariates ($P = 122$) is adequate to adjust for residual confounding. This assumption is not testable, but we argue that it is unlikely that there exists covariates that are uncorrelated with the $P = 122$ observed covariates and that can lead to confounding bias.

We applied inverse probability weighting (IPW)[23–26] to the full cohort and to the low pollution cohort (LPC) to estimate the causal hazard rate ratio, which can be interpreted as the hazard of mortality (or hospitalization) at any time t had all subjects been exposed to $PM_{2.5}$ levels higher than 12 μg/m^3 (in the low pollution cohort: higher than 8 μg/m^3, but always lower than 12 μg/m^3) divided by the hazard of mortality (or hospitalization) at time t had all subjects had been exposed to $PM_{2.5}$ levels lower than 12 μg/m^3 (in the low pollution cohort: lower than 8 μg/m^3). The estimation of causal effects using IPW involves two steps: (1) estimation of the inverse probability weights, denoted sw_i and (2) fitting a Cox proportional hazards model[26] to the observations weighted by sw_i. Specifically:

Step 1: Inverse Probability Weighting: Let T_i represent the binary exposure for subject i. More specifically, we assumed that $T_i = 0$ when $PM_{2.5} < 12$ and $T_i = 1$ when $PM_{2.5} > 12$ for the full cohort. In the low pollution cohort, $T_i = 0$ when $PM_{2.5} < 8$ or $T_i = 1$ when $8 < PM_{2.5} < 12$. We denote by C_i the full set ($P = 122$) of individual level and ZIP code level covariates. For each subject, we estimate sw_i as follows:

$$sw_i = \frac{P\left(T_i = t_i\right)}{P(T_i = t_i | C_i = c_i)}$$

IPW weighting should produce a weighted sample where the distribution of covariates is balanced with respect to T_i, and hence allow a causal estimate of the effect of T_i.

Step 2: Cox Proportional Hazards Model: We then fit to the data a Cox proportional hazards model where every individual observation is weighted by sw_i. The left tail and the right tail of the weights are truncated at the 10th and 90th quantiles of the distribution of the standardized weights, to mitigate the effect of excessively large or small weights.[25,28] Time to event is calculated as the time from reference date until death, the first respiratory, circulatory, or all-cause hospitalization (Figure 2). Death dates are censored at the end of the 1-year outcome observation period. Hospitalization dates are censored at the end of the 1-year outcome observation period or death, whichever comes first. We calculate 95% confidence intervals based on robust, sandwich variance estimators[29] to take into account within-subject correlation induced by repeated measures, the standardized weights, and correlation between subjects living in the same ZIP code.

To measure the potential public health impact of lowering pollution levels below 12 μg/m^3, we calculated the number of events attributable to a change in long-term exposure to $PM_{2.5}$ from below 12 μg/m^3 to above 12 μg/m^3. We used the formula $A = N \times (1 - [1/HR])$, where HR is the hazard ratio comparing exposure above and below 12 μg/m^3, N is the number of events in the Medicare population, and A is the number of events attributable to an increase in $PM_{2.5}$ from below to above 12 μg/m^3.

Sensitivity Analyses

We conducted several sensitivity analyses, summarized in eTable 2; http://links.lww.com/EDE/B215. First, to directly compare our results to the American Cancer Society (ACS) Cohort and the Harvard Six Cities studies,[5,6,30–32] we analyzed the data using a standard Cox proportional hazards model with continuous exposure and adjustment for confounding by including all the available covariates as linear terms into the model (SA1, eFigure 2; http://links.lww.com/EDE/B215 and eTable 3; http://links.lww.com/EDE/B215). Second, we performed a Wald test to assess if there is evidence of the nonlinearity of the exposure–response function (SA2, eTable 4;

http://links.lww.com/EDE/B215), and we plotted the resulting nonlinear exposure–response curves (SA2, eFigure 3; http://links.lww.com/EDE/B215). Third, we ran the analyses restricting to subjects living in areas with long-term exposure to $PM_{2.5}$ less than 12 μg/m^3, though we use as an exposure a binary indicator of being below 10 μg/m^3 instead of 8 μg/m^3 as done in the main analysis (SA3, eFigure 4; http://links.lww.com/EDE/B215 and eTable 5; http://links.lww.com/EDE/B215). Finally, we investigated the sensitivity of the results to the exclusion of the behavioral risk factors extracted from MCBS data (e.g., smoking, BMI) from the confounding adjustment.

RESULTS

Table 1 summarizes the main characteristics of the MCBS–Medicare cohort (for both the full and low pollution cohorts) in comparison to the characteristics of the cohorts from the two original landmark studies—the ACS and Six Cities studies.[5,6,30–32] Note that in our study, the average level of $PM_{2.5}$ (equal to 12 μg/m^3) is substantially lower than what was observed in the Harvard Six Cities Study and in the ACS Cohort (16.4 and 17.7 μg/m^3, respectively).

Figure 3 shows the average $PM_{2.5}$ exposure in the 5138 ZIP codes (1067 unique counties) where MCBS enrollees resided in 2002. During the 1-year follow-up period from the reference date, 4.95% died, 22.2% had one or more hospitalizations, 19% were hospitalized at least once with a circulatory disease, and 9.7% were hospitalized at least once for a respiratory disease.

Table 2 summarizes the results of IPW applied to both the full cohort and the low pollution cohort. We estimated that increasing long-term exposure to $PM_{2.5}$ from levels lower than 12 μg/m^3 to levels higher than 12 μg/m^3 causally increases all-cause admissions and circulatory admission hazard rates by 7% (95% CI = 3%, 10%) and 6% (95% CI = 2%, 9%), respectively. This implies that the total number of all-cause admissions and circulatory admissions from 2002 to 2010 in Medicare attributable to an increase in long-term average $PM_{2.5}$ levels from below 12 μg/m^3 to above 12 μg/m^3 is estimated to be 5,861,028 and 1,417,962, respectively. We did not find evidence of an increase in mortality or respiratory admissions. We also estimated that in the low pollution cohort, increasing $PM_{2.5}$ levels from below 8 μg/m^3 to above 8 μg/m^3 (but always lower than 12 μg/m^3) causally increases all-cause, circulatory, and respiratory admission hazard rates by 15% (95% CI = 8%, 23%), 18% (95% CI = 10%, 27%), and 21% (95% CI = 9%, 34%), respectively, and all these effects were statistically significant. We did not find evidence of an increase in mortality.

Figure 4 illustrates the sensitivity of the results summarized in Table 2 with respect to omission of all the MCBS variables when estimating sw_i. Each panel summarizes the results for a different outcome (all-cause hospitalization, circulatory hospitalization, death, respiratory hospitalization). Within each panel, we illustrated the results for both the full cohort and LPC. Estimates in red are obtained when we use the entire set of all the available potential confounders to adjust for confounding (122 potential confounders). Estimates in blue (claims only) are obtained when we exclude the MCBS variables ($P = 122–73 = 49$) in the approach for confounding adjustment. The fact that blue and red estimates are highly overlapping, indicate that our conclusions are robust to the exclusion of the MCBS variables among the confounding variables used for the adjustment.

More generally, results from the sensitivity analyses (SA1, SA2, and SA3) mentioned in the Methods section and reported in the supplementary material suggest that our estimates are

TABLE 1. Summary Statistics of the MCBS–Medicare Full and Low Pollution Cohorts in Comparison to Other Cohorts

Characteristic	MCBS–Medicare Full Cohort	MCBS–Medicare Low Pollution Cohort (Cohort with Annual $PM_{2.5}$ <12 μg/m^3)	American Cancer Society Cohort (Pope et al)[6,12]	Harvard Six Cities Study Cohort (Dockery et al[5] and Laden et al[31])
No. individuals	32,119	18,144	~293,000	~8,000
Mean age at enrollment	72.0	72.3	58.6	49.7
No. years of follow-up from interview date	1	1	18	24
Study period	2002–2010	2002–2010	1982–2000	1974–1998
Time period where exposure was measured	2000–2010	2000–2010	1979–1983, 1999–2000	1979–1988, 1990–1998
Spatial resolution for exposure assessment	ZIP codes (N = 5,138)	ZIP codes (N = 3,079)	Counties (N = 50)	Cities (N = 6)
$PM_{2.5}$ during the study period (μg/m^3); mean (IQR_W)	12 (3.41)	10.18 (2.46)	17.7 (3.7)	16.4 (5.6)
No. confounders	122	122	~50	~40

Table 1 summarizes the main characteristics of the MCBS–Medicare cohort (for both the full and low pollution cohorts) in comparison to the characteristics of the cohorts from the two original landmark studies–the ACS and Six Cities studies.[5,6,30–32] Note that in our study, the average level of $PM_{2.5}$ (equal to 12 μg/m^3) is substantially lower than what was observed in the Harvard Six Cities Study and in the ACS Cohort (16.4 and 17.7 μg/m^3, respectively).

IQR_w indicates interquartile range width.

TABLE 2. Hazard Ratios Showing the Effect of Living in a High Pollution Versus Low Pollution

	Full Cohort, Threshold = 12 μg/m^3, N = 32,119, Person-years = 68,789	Low Pollution Cohort (Cohort with Annual $PM_{2.5}$ <12 μg/m^3), Threshold = 8 μg/m^3, N = 18,144, Person-years = 34,429
All-cause mortality	0.97 (0.90, 1.04)	1.11 (0.97, 1.28)
All-cause hospitalization	1.07 (1.03, 1.10)	1.15 (1.08, 1.23)
Circulatory hospitalization	1.06 (1.02, 1.09)	1.18 (1.10, 1.27)
Respiratory hospitalization	1.03 (0.98, 1.08)	1.21 (1.09, 1.34)

Hazard ratios are computed using inverse probability weighting. Table 2 reports 95% confidence intervals based on robust, sandwich variance estimators.

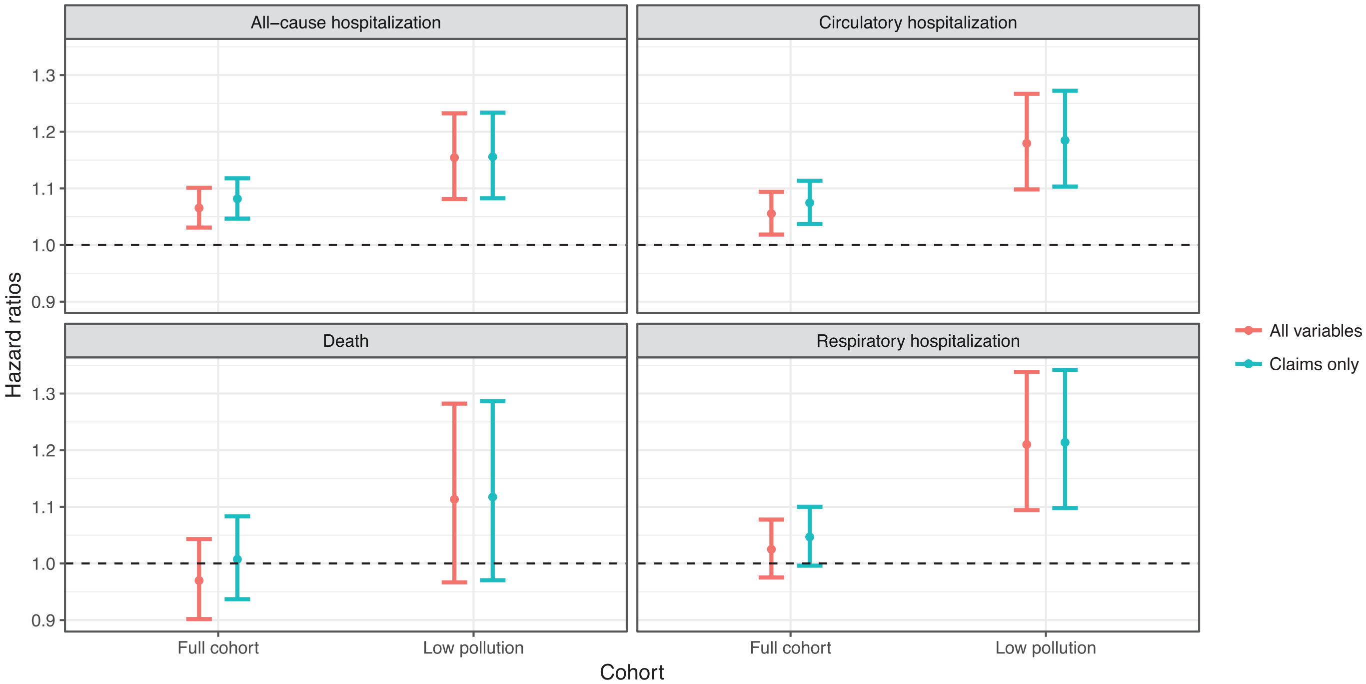

FIGURE 4. Sensitivity to exclusion of MCBS variables: hazard ratios and 95% confidence intervals based on robust, sandwich variance estimators computed including (red) and excluding (blue) MCBS variables.

largely robust across different statistical methodologies, model misspecification, and confounder exclusion. Importantly, as summarized in the supplemental material, our analyses using a standard Cox proportional hazards model with continuous exposure also found significant effects for hospitalizations.

DISCUSSION

We have combined several sources of data and constructed the MCBS–Medicare cohort to address the following three questions: (1) does increasing the level of $PM_{2.5}$ from below 12 μg/m^3 to above 12 μg/m^3 causally increase deaths and hospitalizations; (2) among individuals with exposure levels below 12 μg/m^3, does increasing the level of $PM_{2.5}$ from below 8 μg/m^3 to above 8 μg/m^3 causally increase deaths and hospitalizations; and (3) does exclusion of individual level behavioral risk factors materially affect our estimates?

The Harvard Six Cities Study[5,31] and the ACS Study[6,12] are two landmark epidemiologic cohort studies that had an enormous impact on our understanding of the health effects of air pollution. However, these studies have limited statistical power to detect the effects of low levels of air pollution, particularly because most of their subjects reside in urban areas where pollution levels tend to be higher. The Six Cities Study[5,31] and the ACS Study[6,12] are also limited by the fact that they are closed cohort studies in the sense that they do not allow enrollment of new individuals into the cohort. As such, these studies are less able to estimate the health effects of recent air pollution, nor can they track health effects over time. To overcome this challenge, more recent epidemiologic studies have leveraged "open" cohort data, such as Medicare claims, which permit new enrollees to enter the cohort each year. Our study leverages Medicare claims data combined with data on individual level behavioral risk factors, an important factor missing in previous studies. Including individual level behavioral risk factors in our analysis is very important as these factors are generally hard to measure and are only available from cohort studies. To our knowledge, this is the first epidemiologic study that estimates the effects of low levels of air pollution using claims data augmented with individual level behavioral risk factors, thus overcoming the common criticism that studies that rely entirely on claims data are myopic to important potential confounders.

Our study uses inverse probability weighting, enabling us to estimate causal effects. The results are consistent with existing literature on the adverse health effects of long-term exposure to $PM_{2.5}$. We found robust evidence that increasing long-term exposure to $PM_{2.5}$ (2 years average) from levels lower than 12 μg/m^3 to levels higher than 12 μg/m^3 increases all-cause admissions and circulatory admission hazard rates; and among individuals with exposure levels below 12 μg/m^3, exposure to $PM_{2.5}$ levels above 8 μg/m^3 increases all-cause, circulatory, and respiratory admission hazard rates. We also found evidence that the marginal benefit is increasing at lower concentrations: in the low pollution cohort, an increase of $PM_{2.5}$ from below 8 μg/m^3 to above 8 μg/m^3 led to a 15% increase in hospitalization rate, whereas in the full cohort, an increase of $PM_{2.5}$ from below 12 μg/m^3 to above 12 μg/m^3 led to a 7% increase in hospitalization rate. This evidence is consistent with our previous work.[34] Our additional analyses, which include the whole Medicare population, have relied on much larger statistical power to test this hypothesis.[35]

Our study has several strengths that can be leveraged in future studies. Previous studies assign each subject an average exposure aggregated at the county or at the larger metropolitan area level, which is a coarse indicator of a subject's exposure to air pollution that lends itself to exposure measurement error.[36,37] For this study, we estimate exposure on a 1 × 1 km grid to compute exposure at the ZIP code level. These estimates, obtained from previous work,[20,38–41] allow us to directly study the effects of low levels of pollution with an unprecedented scale of spatial resolution. Importantly, we also investigated the sensitivity of the results when we exclude from the confounding adjustment all of the behavioral risk factors ($P = 73$) measured in the MCBS (e.g., smoking, BMI) and found that the results do not change. This finding indicates that claims data combined with ZIP code level data on risk factors and socioeconomic data are sufficient to rigorously estimate the health effects of air pollution when using ZIP code level exposure data. Thus, results from this study indicated that expensive and potentially time consuming collection of a large set of individual level behavioral risk factors, although potentially useful for exploring susceptibility and effect modification, is not critical to adjust for confounding bias. Furthermore, the results of this analysis add validity to air pollution epidemiologic investigations that rely entirely on administrative and therefore publicly available data.

Despite robustness of results, our results have certain limitations that will be important to address in future studies. Our study population is markedly smaller than the population included in the ACS Study (Table 1). To increase our sample size, we included all individuals who had an MCBS interview at any point during the study period 2002–2010, thus restricting the follow-up period to only 1 year. The limited sample size and limited follow-up period might be the reason why we did not find an effect for mortality, only 4.95% of whom died versus 22.2% who were hospitalized. In another recent study conducted by the same team and that includes the entire Medicare population (approximately 60 million participants) with an average follow-up of 7 years, we report an association between long-term exposure to $PM_{2.5}$ and mortality, even at levels below the 12 μg/m^3.[35] Another limitation in our study was analyzing the data assuming that exposure is binary and time invariant. These are strong assumptions but allow for simple interpretation of the results and for visual inspection of the balance across covariates before and after stratifying on the estimated propensity score, thus substantially increasing the level of confidence in our results with respect to proper adjustment for confounding. In addition, we conducted analyses using a continuous exposure and a Cox proportional hazard model and found the same results.

As more data become available, future studies will be able to repeat these analyses routinely and with a longer follow-up period. In addition, because our cohort is open in the sense that it allows for new enrollment every year (US elderly >65 that enters into fee-for-service Medicare), our findings allow for continued monitoring of the health effects as air pollution continues to decline. Our analyses can be repeated routinely every few years as new claims data become available to track the effectiveness of regulatory actions and mitigation strategies over time. Also, unlike more traditional closed cohort prospective studies, this study utilizes publicly available data, which permits other entities with access to the Medicare claims data to reproduce our results as a validity check.

Results from this study have important implications for policymakers. With data from 5138 unique ZIP codes, spanning 1067 unique counties over a period of 9 years and measuring 122 potential confounders, this work provides compelling evidence that substantial public health benefits have accrued from compliance with the annual NAAQS. The evidence also suggests that further reductions in $PM_{2.5}$ below the current NAAQS would provide additional benefits. The number of cases avoided as a result of compliance is large compared with most public health measures.

ACKNOWLEDGMENTS

We would like to thank Christine Choirat who assisted with data analysis and Leigh Melanson for providing editorial assistance.

REFERENCES

1. Regulatory Impact Analysis for the Final Revisions to the National Ambient Air Quality Standards for Particulate Matter. 474 Leaves in Various Folations: Illustrations (some Color), Charts, Maps, n.d. Web. Available at: https://www3.epa.gov/ttnecas1/regdata/RIAs/finalria.pdf.
2. U.S. Environmental Protection Agency. *Benefits and Costs of the Clean Air Act from 1990 to 2020: Summary Report*. BiblioGov; 2012.
3. Portney PR. Policy watch: economics and the Clean Air Act. *J Econ Perspect*. 1990;4:173–181.
4. US Environmental Protection Agency. *The Benefits and Costs of the Clean Air Act, 1990 to 2020*. 2011.

5. Dockery DW, Pope CA 3rd, Xu X, et al. An association between air pollution and mortality in six U.S. cities. *N Engl J Med*. 1993;329:1753–1759.
6. Pope CA 3rd, Burnett RT, Thun MJ, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA*. 2002;287:1132–1141.
7. Samet JM, Zeger SL, Dominici F, et al. The National Morbidity, Mortality, and Air Pollution Study. Part II: morbidity and mortality from air pollution in the United States. *Res Rep Health Eff Inst*. 2000;94(pt 2):5–79.
8. Samet JM, Dominici F, Curriero FC, Coursac I, Zeger SL. Fine particulate air pollution and mortality in 20 U.S. cities, 1987-1994. *N Engl J Med*. 2000;343:1742–1749.
9. Dominici F, Peng RD, Bell ML, et al. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA*. 2006;295:1127–1134.
10. Dominici F, Sheppard L, Clyde M. Health effects of air pollution: a statistical review. *Int Stat Rev*. 2003;71:243–276.
11. Dockery DW, Pope CA 3rd. Acute respiratory effects of particulate air pollution. *Annu Rev Public Health*. 1994;15:107–132.
12. Pope CA 3rd, Thun MJ, Namboodiri MM, et al. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *Am J Respir Crit Care Med*. 1995;151(3 pt 1):669–674.
13. Breysse PN, Delfino RJ, Dominici F, et al. US EPA particulate matter research centers: summary of research results for 2005–2011. *Air Qual Atmos Health*. 2013;6:333–355.
14. Vedal S, Brauer M, White R, Petkau J. Air pollution and daily mortality in a city with low levels of pollution. *Environ Health Perspect*. 2003;111:45–52.
15. Schwartz J, Bind MA, Koutrakis P. Estimating causal effects of local air pollution on daily deaths: effect of low levels. *Environ Health Perspect*. 2017;125:23–29.
16. DeVries R, Kriebel D, Sama S. Low level air pollution and exacerbation of existing copd: a case crossover analysis. *Environ Health*. 2016;15:98.
17. Crouse DL, Peters PA, Hystad P, et al. Ambient PM2.5, O3, and NO2 exposures and associations with mortality over 16 years of follow-up in the Canadian Census Health and Environment Cohort (CanCHEC). *Environ Health Perspect*. 2015;123:1180–1186.
18. Medicare Current Beneficiary Survey [Da - Centers for Medicare & Medicaid Services. N.p., n.d. Web. Available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/. Accessed 16 March 2016.
19. Burnett RT, Pope AC 3rd, Ezzati M, et al. An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environ Health Perspect*. 2014;122:397.
20. Di Q, Kloog I, Koutrakis P, Lyapustin A, Wang Y, Schwartz J. Assessing PM2.5 exposures with high spatiotemporal resolution across the continental United States. *Environ Sci Technol*. 2016;50:4712–4721.
21. Census.gov. Census.gov. N.p., n.d. Web. Available at: www.census.gov. Accessed 23 January 2016.
22. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health*. 2000;21:121–145.
23. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*. 2000;11:561–570.
24. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550–560.
25. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168:656–664.
26. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006;60:578–586.
27. Cox DR, Oakes D. *Analysis of Survival Data. Book 21*. Boca Raton, FL: CRC Press; 1984.
28. Lee BK, Lessler J, Stuart EA. Weight trimming and propensity score weighting. *PLoS One*. 2011;6:e18174.
29. Liang K, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73:13–22.
30. Pope CA 3rd, Dockery DW. Health effects of fine particulate air pollution: lines that connect. *J Air Waste Manag Assoc*. 2006;56:709–742.
31. Laden F, Schwartz J, Speizer FE, Dockery DW. Reduction in fine particulate air pollution and mortality: extended follow-up of the Harvard Six Cities Study. *Am J Respir Crit Care Med*. 2006;173:667–672.
32. Health Effects Institute. *Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality: A Special Report of the Institute's Particle Epidemiology Reanalysis Project*. Cambridge MA: Health Effects Institute; 2000.
33. Samet JM. The Clean Air Act and health—a clearer view from 2011. *N Engl J Med*. 2011;365:198–201.
34. Shi L, Zanobetti A, Kloog I, et al. Low-concentration PM2.5 and mortality: estimating acute and chronic effects in a population-based study. *Environ Health Perspect*. 2016;124:46–52.
35. Di Q, Wang Y, Zanobetti A, et al. Air pollution and mortality in the entire Medicare population. *N Engl J Med*. 2017: In press.
36. Alexeeff SE, Schwartz J, Kloog I, Chudnovsky A, Koutrakis P, Coull BA. Consequences of kriging and land use regression for PM2.5 predictions in epidemiologic analyses: insights into spatial variability using high-resolution satellite data. *J Expo Sci Environ Epidemiol*. 2015;25:138–144.
37. Gryparis A, Paciorek CJ, Zeka A, Schwartz J, Coull BA. Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics*. 2009;10:258–274.
38. Kloog I, Koutrakis P, Coull BA, Lee HJ, Schwartz J. Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmos Environ*. 2011;45:6267–6275.
39. Kloog I, Chudnovsky A, Just AC, et al. A new hybrid spatio-temporal model for estimating daily multi-year PM2.5 concentrations across northeastern USA using high resolution aerosol optical depth data. *Atmos Environ*. 2014;95:581–590.
40. Lee M, Kloog I, Chudnovsky A, et al. Spatiotemporal prediction of fine particulate matter using high-resolution satellite images in the Southeastern US 2003-2011. *J Expo Sci Environ Epidemiol*. 2016;26:377–384.
41. Kloog I, Nordio F, Coull BA, Schwartz J. Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM2.5 exposures in the Mid-Atlantic states. *Environ Sci Technol*. 2012;46:11913–11921.