Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/sciadv.aba5692/DC1

Supplementary Materials for

Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly

X. Wu, D. Braun, J. Schwartz, M. A. Kioumourtzoglou, F. Dominici*

*Corresponding author. Email: fdominic@hsph.harvard.edu

Published 26 June 2020, *Sci. Adv.* **6**, eaba5692 (2020)
DOI: 10.1126/sciadv.aba5692

This PDF file includes:

S.1 Statistical Methods

S.1.1 Notations

We fit five different statistical models to estimate the causal relationship between long-term $PM_{2.5}$ exposure and our outcome of interest, all-cause mortality among the elderly. We use the following mathematical notation: let E indicate the continuous $PM_{2.5}$ exposure ranging from e_0 to e_1; let $\boldsymbol{X}$ indicate the p-dimensional vector of measured potential confounders; let Y indicate the health outcome (here all-cause mortality); let $\boldsymbol{C}$ denote the individual-level stratifying variables; and let N indicate the sample size, with $j = 1, ..., N$ indexing Medicare enrollees in the sample.

S.1.2 Cox Proportional Hazard Model

Studies investigating the association between long-term exposure to $PM_{2.5}$ and mortality have traditionally applied the Cox proportional hazards model (1, 3), a commonly-used approach for survival analysis. We fit the following stratified Anderson-Gill Cox proportional hazard model with follow-up times a as the time metric (34):

$$h^{\boldsymbol{c},z}(a,t) = h_0^{\boldsymbol{c}}(a) exp(\beta_1 E_{z,t} + \boldsymbol{\beta_2 X}_{z,t}) \tag{1}$$

where $h^{\boldsymbol{c},z}(a,t)$ denotes the hazard for mortality for individuals who were in strata $\boldsymbol{c}$, resided in zip code z at follow-up year a and calendar year t; and $h_0^{\boldsymbol{c}}(a)$ is a strata-specific baseline hazard function. $E_{z,t}$ is the exposure (i.e. annual average $PM_{2.5}$ concentrations) at calendar year t in zip code z. To adjust for confounding bias, we included $\boldsymbol{X}_{z,t}$, the **fourteen** zip code and county-level time-varying covariates at calendar year t in zip code z, in the model. We adjusted for potential residual spatial and temporal confounding by including a dummy spatial variable (census region) and a dummy temporal variable (calendar year). Assuming a constant baseline hazard within each follow-up year, this Cox model uses follow-up year a as the time metric, and thus creates a piece-wise exponential hazard for individuals with the same follow-up year. In addition, to adjust for potential confounding bias by individual-level characteristics and handle the potential non-proportionality of individual hazards, a different baseline hazard function was specified for each stratum defined by the **four** individual-characteristics.

S.1.3 Poisson Regression Model

We also fit the following Poisson regression model

$$log\mathbb{E}[Y_{a,t}^{c,z}] = log(T_{a,t}^{c,z}) + log(h_0^c(a)) + \beta_1 E_{z,t} + \boldsymbol{\beta_2 X}_{z,t} \tag{2}$$

where $Y_{a,t}^{c,z}$ is the total death count for enrollees who were in each stratum $\boldsymbol{c}$, and resided in zip code z at follow-up year a and calendar year t; and $T_{a,t}^{c,z}$ is the corresponding total person-time. There is a wide literature showing the mathematical equivalency of the Cox proportional hazard model and this specific formulation of the Poisson regression model (35–37).

S.1.4 Potential Outcome and Generalized Propensity Score Framework

We considered three causal inference modeling approaches based on 1) matching by GPS (38); 2) weighting by GPS (30); and 3) including GPS as a covariate in the health outcome model (adjustment by GPS) (31). We begin by describing the general framework for causal inference.

A key concept of causal inference is potential outcomes, sometimes referred to as counterfactual outcomes. The potential outcomes framework was first proposed by Neyman in 1923 (39) in the context of fully randomized experiments. Rubin, together with other contemporary statisticians, extended this framework into a general framework for thinking about causation in both observational and experimental studies (40). Briefly, a potential outcome is the outcome that would have been realized if an individual had received a specific value of the exposure.

Definition 1 *The potential outcomes are defined as a set of random variables,* $Y(e), \forall\ e \in [e^0, e^1]$*, in which* $Y_j = Y_j(E_j), \forall\ E_j \in [e^0, e^1], \forall\ j = 1, ..., N.$

The potential outcomes framework, although not the only approach used to frame and answer questions about causality (41), is very appealing and convenient both for the sake of logical completeness (30) and for answering real-world problems (42; 43).

Using propensity scores (PS) to adjust for confounding in a potential outcome framework is one, very common, approach for studying causal effects in observational studies (44) (this seminal paper has received more than 25,000 citations). We begin by defining the standard PS, which

requires a binary exposure.

Definition 2 *For $E \in \{0, 1\}$, propensity scores (PS) is the conditional probability of receiving the exposure given potential confounders:* $q(\mathbf{x}) = Pr(E = 1 \mid X = x)$.

A key property of the PS is called the balancing property; conditional on the same propensity score value, the probability of receiving an exposure is independent of X (44). It allows one to simultaneously balance a large set of covariates in the exposed $(E = 1)$ and reference populations $(E = 0)$. By ensuring covariate balance between the exposed population and a reference population a pseudo-population is created which mimicks a randomized experiment (45). Randomized experiments are considered the ”gold standard” to inform causality (46–48) and ensure the covariate distributions do not differ by exposure status, that is, the covariates are balanced. These randomized experiments achieve covariate balance between exposed and reference populations with respect to both measured and unmeasured covariates, whereas the pseudo-population created by using PS approaches in observational studies can achieve covariate balance with respect to the measured covariates.

However the use of standard PS approaches requires a binary exposure, which is often not the case in the majority of environmental health research where the exposure is continuous. Approaches to estimate causal exposure-response curves (ERCs) have been proposed, including methods that rely on the generalized propensity score (GPS) (38; 49). The GPS is an analogue to the PS for continuous exposures and also satisfies a balancing property described in (31).

Definition 3 *The GPS is the conditional density function of the exposure given potential confounders :* $\mathbf{q}(\mathbf{x}) = \{f_{E|\mathbf{X}}(e \mid \mathbf{x}), \forall e \in [e^0, e^1]\}$. *The single score* $q(e, \mathbf{x}) = f_{E|\mathbf{X}}(e \mid \mathbf{x})$ *are called realizations of* $\mathbf{q}(\mathbf{x})$ *at exposure level* e.

The potential outcome and GPS framework provide tools to estimate the causal estimand and discuss modeling assumptions. We follow the large body of literature in causal inference to state the following assumptions of identification.

Assumption 1 (Consistency) $E = e$ *implies* $Y = Y(e)$.

This assumption, also refered to as no-interference (50), or the stable-unit-treatment-value as-

sumption (SUTVA) (47). In brief, we assume that the potential outcome for a given observation is not affected by the exposure of any other unit, and that each exposure defines a unique outcome for each observation.

Assumption 2 (Overlap) *For all values of potential confounders* $\mathbf{x}$*, the density function of receiving any possible exposure level* $e \in [e^0, e^1]$ *is positive:* $f(e \mid \mathbf{x}) > 0$ *for all* $e,\ \mathbf{x}$.

This assumption, sometimes referred to as the positivity assumption, states that the exposure is not assigned deterministically, and thus each individual has a positive chance of receiving any exposure level e, regardless of potential confounders $\mathbf{X}$. It guarantees that for all possible values of potential confounders $\mathbf{x}$, we will be able to estimate $\mu(e)$ for each exposure level e without relying on extrapolation.

Assumption 3 (Weak Unconfoundedness) *For any possible exposure level* e*, in which* e *is continuous in the range* $[e^0, e^1]$*;* $E \perp\!\!\!\perp Y(e) \mid \mathbf{X}$.

This assumption, sometimes referred to as the ignorability assumption, states that the mean potential outcome under level e is the same across treatment levels once we condition on potential confounders (i.e. exposure assignment is unrelated to potential outcomes within strata created by potential confounders). This assumption indicates the possibility that if sufficiently many relevant covariates $\mathbf{X}$ are collected, we would be able to approximate a stratified randomized experiment from observational studies by conditioning on the set of covariates $\mathbf{X}$.

The three causal assumptions stated above allow us to identify and estimate the following causal estimand; the average causal ERC (38; 49).

Definition 4 *The average causal ERC is* $\mu(e) = \mathbb{E}[Y(e)]$*, for all* $e \in [e^0, e^1]$.

Other causal estimands that are constructed directly based on average causal ERCs can then be causally identified and estimated, including various types of ratio quantities (51).

S.1.5 Causal Inference Approaches

The main advantage of causal inference approaches compared to more traditional approaches is that their ”design” and ”analysis” stages are separate (52; 53). In the design stage, investigators

design the study creating a pseudo-population which mimicks a randomized experiment, without using the outcome information. Only after the ”design” stage is complete does the ”analysis” stage begin, conducting outcome analysis on the pseudo-population. In practice, all approaches that rely on GPS include four steps; 1) estimation of the GPS, where exposure is regressed on the potential confounders, 2) implementation of the GPS model to create a pesudo-population, 3) assessing the quality of the constructed pesudo-population in terms of covariate balance, 4) outcome model analysis on the pseudo-population if the pseudo-population is balanced (46). Steps 1-3 belong to ”design” stage and step 4 belongs to the ”analysis” stage. An overview of the workflow for implementing causal inference approaches using GPS to design and analyze observational data is presented in Figure S1.

In the first step, we estimate the GPS. As defined in Section **S.**1.4, the GPS is a density function and thus is estimated by density estimation approaches. Various flexible parametric/non-parametric density estimation models are proposed in (49).

In the second step, referred to as GPS implementation, we use the estimated GPS to adjust for confounding bias. We consider the following three common and validated causal inference approaches; 1) matching by GPS (38); 2) weighting by GPS (30); and 3) including GPS as a covariate in the health outcome model (adjustment by GPS) (31).

S.1.5.1 Matching Approach

Following Wu et al. (38), we implement a nearest-neighbor matching algorithm. Like all other matching algorithms, it starts with defining a matching function with a specified distance measure. Wu et al. (38) proposed a two-dimensional matching function that calculates the joint distance of exposures and GPS with scaled Mahalanobis distance, and tries to find the matched pairs with minimal total distance. The matching function also contains a scale parameter λ to control the relative weights of two dimensions and caliper δ to forbid large deviations (54). The procedure for the matching approach is as follows;

1. Define a suitable caliper matching function with Mahalanobis distance by specifying the scale parameter $\lambda = 1$, and caliper $\delta = 0.24$ as the interval width between fifty equidistant counter-

factual levels $e \in [e^0, e^1]$. The selection of tuning parameters δ follows a data-driven procedure to achieve the best covariate balance in terms of AC, and we conducted a grid search with the following number of counterfactual levels $(25, 50, 100, 200)$ where 50 was selected to achieve optimal covariate balance.

2. Match individuals based on the specified caliper metric matching function. Impute $Y_j(w)$ as: $\hat{Y}_j(e) = Y^{obs}_{m_{\text{GPS}}(e,q(e,\mathbf{x}_j))}$ for each individual $j = 1, \ldots, N$ successively. Construct the matched pseudo-population by imputing $\hat{Y}_j(e)$ for each levels of the exposure $e \in [e^0, e^1]$. Matching is only allowed within each strata defined by the same **four** individual-level characteristics, and within the same followup year.

S.1.5.2 Weighting Approach

Following Robins et al. (30), the weighting approach is involves using the inverse of (generalized) propensity score to weigh the observations. "Stabilizing" the weights is often advised in practice to help reduce estimation variance especially when the exposure is continuous. In order to stabilize the weights we multiply the inverse of the GPS by the marginal probabilities of exposure. A trimming technique is also proposed to avoid extremely large weights (55). The procedure for the GPS weighting approach is as follows;

1. Compute a stabilized version of the inverse GPS weight using the inverse of the estimated GPS, that is $W_{stable} = \int_{\S} \hat{q}(e, \mathbf{x}) d\mathbf{x} / \hat{q}(e, \mathbf{x})$. Trim extreme weight values by setting all weights greater than 10 to 10 (46; 56).

2. Assign weights to each individual $j = 1, ..., N$ to create a weighted pseudo-population.

S.1.5.3 Adjustment Approach

Following Hirano and Imbens (31), a covariate adjustment approach includes the estimated GPS $\hat{q}(e, \mathbf{x})$ as a covariate in outcome model. Hirano and Imbens (31) show that including the estimated GPS as a covariate together with the exposure in a bivariate outcome model can remove confounding bias when estimating the causal ERCs. The procedure is as follows;

1. Model the conditional expectation of outcome Y given exposure E and the estimated GPS,

$\hat{q}(E, \mathbf{X})$, as a Poisson regression with flexible formulation of a bivariate function, $\mathbb{E}(Y \mid E, \hat{q}(E, \mathbf{X})) = \mu_{\boldsymbol{\beta}}^{-1}(E = e, \hat{q}(E, \mathbf{X}) = \hat{q}(e, \mathbf{x}))$, where $\mu_{\boldsymbol{\beta}}$ is a link function.

2. Given the estimated parameters, $\hat{\boldsymbol{\beta}}$, from the stratified Poisson regression, we obtain the counterfactual hazard rate as the response variable, $\mu_{\hat{\boldsymbol{\beta}}}^{-1}\{\text{PM}_{2.5}, \text{GPS})\}$ = $\mathbb{E}$(death counts)/person year (57). We then impute the average causal ERC as, $\hat{\mathbb{E}}\{Y(e)\} = \frac{1}{N}\sum_{j=1}^{N} \mu_{\hat{\boldsymbol{\beta}}}^{-1}\{e, GPS_j)\}$, at fifty equidistant counterfactual levels $e \in [e^0, e^1]$. Fifty was selected to match the parameter obtained from the grid search described in Section **S.**1.5.1.

S.1.6 Covariate Balance

In the third step, we assess the quality of our study design, and in particular, evaluate the covariate balance for the constructed pseudo-population via absolute correlation. Our balance diagnostics are motivated by the balancing property of the GPS. The key is that if two variables are independent of one another then the correlation between these two variables will be zero. The evaluation of covariate balance via absolute correlation (AC) is proposed in (38; 58).

Formally, we define the pseudo-population created by each of the GPS implementations using the following notation. Let N_{pseudo} indicate the sample size of the pseudo-population, let $E_{i,pseudo}$ indicate the exposure in the pseudo-population for unit i, let $\mathbf{X}_{i,pseudo}$ indicate the p-dimensional set of potential confounders in the pseudo-population for unit i and let $Y_{i,pseudo}$ indicate the outcome in the pseudo-population for unit i. We centralize and orthogonalize the covariates $\mathbf{X}_{i,pseudo}$ and the exposure $E_{i,pseudo}$ as

$$\mathbf{X}^*_{i,pseudo} = \mathbf{S}_{\mathbf{X}}^{-1/2}(\mathbf{X}_{i,pseudo} - \bar{\mathbf{X}}_{i,pseudo}),\ E^*_{i,pseudo} = \mathbf{S}_{E}^{-1/2}(E_{i,pseudo} - \bar{E}_{i,pseudo}),$$

where $\bar{\mathbf{X}}_{i,pseudo} = \sum_{i=1}^{N_{pseudo}} \mathbf{X}_{i,pseudo}/(N_{pseudo})$, $S_{\mathbf{X}} = \sum_{i=1}^{N_{pseudo}}(\mathbf{X}_{i,pseudo} - \bar{\mathbf{X}}_{i,pseudo})(\mathbf{X}_{i,pseudo} - \bar{\mathbf{X}}_{i,pseudo})^T/(N_{pseudo} - 1)$ and $\bar{E}_{i,pseudo} = \sum_{i=1}^{N_{pseudo}} E_{i,pseudo}/(N_{pseudo})$, $\mathbf{S}_E = \sum_{i=1}^{N_{pseudo}}(E_{i,pseudo} - \bar{E}_{i,pseudo})(E_{i,pseudo} - \bar{E}_{i,pseudo})^T/(N_{pseudo} - 1)$.

Based on the balancing condition, in a balanced population, the correlations (e.g., the Pearson's correlation coefficient r) between the exposure and potential confounders should be equal to zero,

that is $\mathbb{E}[\mathbf{X}^*_{i,pseudo}E^*_{i,pseudo}] = \mathbf{0}$. We assess covariate balance in the pseudo population as

$$\Big| \sum_{i=1}^{N_{pseudo}} \mathbf{X}^*_{i,pseudo}E^*_{i,pseudo}\Big| < \boldsymbol{\epsilon}_1,$$

The average ACs are defined as the average of the ACs among all p covariates.

Definition 5 *The average AC is defined as* $\big|\big| \sum_{i=1}^{N_{pseudo}} \mathbf{X}^*_{i,pseudo}E^*_{i,pseudo}\big|\big|_1/p$.

We assess covariate balance after implementing the weighting approach to create a weighted pseudo-population, and implementing the matching approach to create a matched pseudo-population by calculating average ACs in the corresponding pseudo-populations (38; 58). Note, although the GPS adjustment approach is a very popular (generalized) PS approach, there is no transparent way to evaluate covariate balance after implementing this approach, thus covariate balance is not assessed for this approach (46). To avoid potential violation of the overlap assumption due to the inclusion of outlier exposure estimates, we trim extreme weights above 10 when implementing weighting (59), and exclude data with the highest 1 % and lowest 1 % exposure when implementing matching (4).

We further illustrate the relationship between the proposed AC and the standardized mean difference (SMD), a commonly used covariate balance measure in causal inference (46). Under the binary exposure (e.g., $E = 0, 1$) setting, there is a mathematical equation relating the SMD d to the correlation coefficient r (60). For each potential confounder X, we define the two quantities respectively as;

$$d = \frac{\bar{X}_{E=1} - \bar{X}_{E=0}}{SD_{pooled}}, r = Corr(X, E),$$

where $\bar{X}_{E=1}$ and $\bar{X}_{E=0}$ are the mean of potential confounder X in group with $E = 1$ and with $E = 0$ respectively, and SD_{pooled} is the pooled standard deviation for two groups. The following equation holds

$$d = \frac{2r}{\sqrt{1 - r^2}},$$

where d is monotonically increasing with respect to r. Note the "monotonicity" only hold when the exposure is naturally binary, and does not necessary hold when comparing correlation coefficients based on continuous exposures and the associated SMD calculated based on discontimized binary exposures (61)). When $r = 0.1$, $d \approx 0.2$ which matches the cutoff value 0.2 for the SMD suggested in the binary exposure causal inference literature (46; 62). Zhu et al. (29) also provided a heuristic proof for the cut-off value 0.1 for an AC based measure z when the exposure is continuous and link it to the usual cutoff value 0.2 for the SMD in the binary exposure case. Although the proposed measure in Zhu et al. is a Fisher transformed correlation coefficient, that is

$$z = \frac{1}{2} ln\Big(\frac{1+r}{1-r}\Big),$$

whereas r and z are approximately equal when $\mid r \mid \leq 0.5$, which is true for all potential confounders $\mathbf{X}$ in our study. Following the guidance proposed by Zhu et al., we set the cut-off point for good covariate balance as 0.1 for the average AC.

In the fourth (and final) step, we conduct the outcome analysis. Details on the outcome analysis for each of these approaches is described in the Materials and Methods section.

S.2 Additional Analysis Results

2.1% of the Medicare enrollees (corresponding to 4,587 zip codes) were not linked to confounder data and were, thus, not included in analyses. Given this is such a small proportion, we do not expect this exclusion to impact the results. Although this information is not available in the Medicare claims, it is likely that the zip codes (and subsequently enrollees associated with these zip codes) that were not included in analyses were not standard zip codes and for this reason we were not able to link them to data from the US Census, ACS, and the BFRSS at ZCTA. Specifically, zip codes that only serve PO Boxes do not have a corresponding ZCTA. Thus, we assume that some zip codes not linked to ZCTA are likely PO Box-only zip codes. Please note, that information on enrollees with private PO Boxes with a standard zip code attached to them are still included in the analyses. We have compared the characteristics between enrollees included vs. not included in our analysis. Overall, these two populations were quite similar with no meaningful

differences, except the included population had a slightly higher proportion of Whites and Medicaid eligible.

All five statistical approaches were fit on four cohorts; 1) all Medicare enrollees among years 2000-2012, 2) Medicare enrollees who were continuously exposed to low level $PM_{2.5}$ among years 2000-2012, 3) all Medicare enrollees among years 2000-2016, 4) Medicare enrollees who were continuously exposed to low level $PM_{2.5}$ among years 2000-2016. Analyses on the 2000-2012 cohort were conducted as a comparison to previously reported estimates in (63). To evaluate the model sensitivity to some potential unmeasured confounders that vary over time as the exposure and the outcome and that are invariant over locations, all five approaches were fit twice, once with year as a covariate and once without. Additional sensitivity analyses were conducted by fitting models without meteorological variables as covariates.

The Medicare enrollees 2000–2012 cohort consisted of 56,095,877 subjects (415,551,432 person-years); we observed 20,303,529 deaths (36.2%) (Table S2). Figure S2 upper panels present the ACs in this cohort for each covariate in the weighted (blue line), matched population (green line), and unadjusted observational population (red line) when year was included in the GPS model. Using the causal inference GPS approaches (matching and weighting) we achieved excellent balance across potential confounders, mimicking randomized control studies, and strengthening the interpretability and validity of our analyses as providing evidence of causality.

Effect estimates are presented as Hazard Ratio (HR) per $10\ \mu g/m^3$ increase in annual $PM_{2.5}$ 95%, confidence intervals (CI)s for all models were evaluated by m-out-n bootstrap using zip code clusters to account for within zip code spatial correlation (500 replicates). We re-calculate the GPS and outcome model in each bootstrapped sample to ensure the bootstrapping procedures jointly account for the variability associated with the estimation of the GPS and outcome model. For the traditional approaches, all confounders included in the health models were statistically significant for all models and all cohorts.

Our findings across all approaches for the 2000-2012 cohort are consistent and statistically significant - a $10\ \mu g/m^3$ increase in $PM_{2.5}$ leads to an increase in mortality risk ranging between 5 and 7% (HR estimates 1.05–1.07). These findings are robust across all statistical approaches (lower

panels of Figure S2). The estimated HRs were generally larger (1.23 to 1.37) when studying the cohort of Medicare enrollees who were always exposed to $PM_{2.5}$ level lower than $12\ \mu g/m^3$. Di et al. (63) reported a HR of 1.07(95 % CI:$1.07, 1.08$) in a previous association study, which is consistent with our findings. All corresponding numbers are provided in Table S3. The HR estimates are close to those based on 2000–2016 cohort (reported in main text).

In addition to the ACs, we calculated the standardized mean difference (SMD) after dichotomizing the continuous exposure for the 2000-2016 cohort. We consider two meaningful cut-off values for long-term exposure to $PM_{2.5}$; a) $12\ \mu$g/m^3 and b) $10\ \mu$g/m^3, corresponding to the current US standards and the WHO guidelines respectively. We dichotomized the exposure levels using the $12\ \mu$g/m^3 cut-off in the main analysis using all Medicare enrollees among years 2000-2016 and using the $10\ \mu$g/m^3 cut-off in the low-level analysis using all Medicare enrollees among years 2000-2016. Figure S3 shows covariate balance for the newly-defined binary exposures. We found that covariate balance was significantly improved after weighting or matching based on the SMD measure as well.

Figure S4 shows the ACs under the sensitivity analyses for the 2000-2016 cohort in which we exclude year (upper panels of Figure S4) and under which we exclude meteorological variables (lower panels of Figure S4) as confounders in the GPS model. We find that excluding year and meteorological covariates in the GPS model results in an imbalance of these covariates, and thus reduces the credibility of the health analyses results under these settings.

Figure S5 presents the distributions of the GPS estimated by using 1) all Medicare enrollees among years 2000-2012, 2) Medicare enrollees who were continuously exposed to low level $PM_{2.5}$ among years 2000-2012, 3) all Medicare enrollees among years 2000-2016, 4) Medicare enrollees who continuously exposed to low level $PM_{2.5}$ among years 2000-2016. GPS ranges from $[0.00, 0.39]$ for all cohorts.

Table S4 presents the importance scores of each of the variables included in the GPS model for each cohort respectively. The importance scores represent the fractional contribution of each variable to the model based on the total gains (28). We find in the GPS model for cohort 2000-2016, the variables year, population density, summer temperature and summer relative humidity

were given the highest importance scores.

S.3 Additional Sensitivity Analysis: E-Value

We included indicator years to adjust for some unmeasured confounders that vary temporally with the exposure and the outcome, and that are invariant spatially; and indicator census geographic regions to adjust for some unmeasured confounders that vary spatially with the exposure and the outcome, and that are invariant temporally. However, even after adjustment for these indicators, the conclusion of our study could be affected by confounding bias by unmeasured factors. We conduct a sensitivity analysis to unmeasured confounding by calculating the E-value (10, 11). The E-value for the point estimates of interest (in our case the hazard ratio, HR) can be defined as the minimum strength of an association, on the risk ratio scale, that an unmeasured confounder would need to have with *both* the exposure and outcome, conditional on the covariates already included in the model, to fully explain the observed association under the null. We calculate the E-values for our reported HRs per 10 μg/m^3 increase of long-term exposure to $PM_{2.5}$.

Table S5 summarizes the results for this sensitivity analysis. For example, for our main analysis (2000–2016) under a Poisson model, we found that for an unmeasured confounder U to fully account for the estimated effects of the exposure E on the outcome Y it would have to be associated with *both* long-term $PM_{2.5}$ exposure (E) and with mortality (Y) by a risk ratio of at least 1.32-fold each, through pathways independent of all covariates already included in the model. In other words, if we were to include this U the association between the long term effects of $PM_{2.5}$ on mortality would become null. A 1.32 risk ratio means that U would need to meet the following two criteria: 1) U would need to lead to a 32% increase in the risk of mortality (Y); *and* 2) when comparing two groups one with exposure to $PM_{2.5}$ that is 10 μg/m^3 higher than the other ($E =$ low versus $E =$ high), the higher exposure group would have a 32% higher prevalence of that unmeasured confounder than the lower exposure group. In our analysis assessing effect estimates of low $PM_{2.5}$ concentrations, for an unmeasured confounder to fully account for the observed results it would have to be associated with both long-term $PM_{2.5}$ exposure and the mortality by a risk ratio of at least 1.76-fold each. The estimated E-values for the low-level analyses were always higher

than the E-values for the main analyses, indicating that the results from the low-levels analyses are less sensitive to unmeasured confounding.

The estimated E-value is conditional on the set of the covariates that we have already included in the model (10). As suggested by VanderWeele and Ding (10), we also calculated the analogues E-value omitting from analysis each of the following key covariates: calendar year and meteorological variables (Table S5). We found the analogues E-values for the calendar year and meteorological variables are smaller than the reported E-values for all three causal inference approaches in our main analysis; the analogues E-values are smaller than the reported E-values for all five approaches in our low level analysis as well. These results suggest our conclusions are robust to unmeasured confounding that would be as strong as the confounding bias caused by calendar year or meteorological variables.

S.4 Code

We provide code for all analyses reported in this paper, including code for the five proposed approaches and code to assess covariate balance. The completed code can be found on `https://github.com/wxwx1993/National_Causal`. Data used in the analyses are stored on a secure cluster hosted by the Research Computing Environment supported by the Institute for Quantitative Social Science in the Faculty of Arts and Sciences at Harvard University.

S.4.1 Code for Loading Required Packages

```
library("dplyr")
library("data.table")
library("fst")
library("survival")
library("gnm")
library("mgcv")
library("xgboost")
library("parallel")
```

S.4.2 Code for Fitting Cox Proportional Hazard Regression

```
# Cox PH
Cox<-coxph(Surv(followup_year,followup_year_plus_one,dead)~pm25_ensemble +
             mean_bmi + smoke_rate + hispanic + pct_blk +
             medhouseholdincome + medianhousevalue +
             poverty + education + popdensity + pct_owner_occ +
             summer_tmmx + winter_tmmx + summer_rmax + winter_rmax +
             as.factor(year) + as.factor(region) +
             strata(as.factor(entry_age_break))+strata(as.factor(sex))+
             strata(as.factor(race))+strata(as.factor(dual)),
           data=national_merged2016,
           ties = c("efron"),na.action = na.omit)
```

S.4.3 Code for Fitting Conditional Poisson Regression

```
# Cox-equvalent conditional Poisson regression
## Create aggregated data for Poisson regression
national_merged2016$time_count<-national_merged2016$followup_year_plus_one -
national_merged2016$followup_year
dead_personyear<-aggregate(cbind(national_merged2016$dead,
    national_merged2016$time_count),
by=list(national_merged2016$zip, national_merged2016$year,
    national_merged2016$sex, national_merged2016$race,
    national_merged2016$dual, national_merged2016$entry_age_break,
    national_merged2016$followup_year), FUN=sum)
confounders<-aggregate(national_merged2016[,c(12:27)],
by=list(national_merged2016$zip, national_merged2016$year,
        national_merged2016$sex, national_merged2016$race,
        national_merged2016$dual,
        national_merged2016$entry_age_break,
```

```
        national_merged2016$followup_year), FUN=min)
aggregate_data<-merge(dead_personyear, confounders,
        by=c("Group.1", "Group.2", "Group.3", "Group.4", "Group.5",
            "Group.6", "Group.7"))
colnames(aggregate_data)[8:9]<-c("dead","time_count")
colnames(aggregate_data)[1:7]<-c("zip","year","sex","race","dual",
                                 "entry_age_break","followup_year")
aggregate_data<-subset(aggregate_data[complete.cases(aggregate_data) ,])
aggregate_data<-merge(aggregate_data,covariates,by=c("zip","year"),all.x=T)

gnm<-gnm(dead ~ pm25_ensemble +
               mean_bmi + smoke_rate + hispanic + pct_blk +
               medhouseholdincome + medianhousevalue +
               poverty + education + popdensity + pct_owner_occ +
               summer_tmmx + winter_tmmx + summer_rmax + winter_rmax +
               as.factor(year) + as.factor(region) +
               offset(log(time_count)), eliminate = (as.factor(sex):
               as.factor(race):as.factor(dual):as.factor(entry_age_break):
               as.factor(followup_year)), data=aggregate_data,
               family=poisson(link="log"))
```

S.4.4 Code for the Estimation of the GPS

```
# The model to estimate GPS
covariates<-aggregate(national_merged2016[,c(12:27)],
        by=list(national_merged2016$zip, national_merged2016$year), FUN=min)
colnames(covariates)[1:2]<-c("zip","year")
covariates<-subset(covariates[complete.cases(covariates) ,])
covariates$year_fac <- as.factor(covariates$year)
covariates$region <- as.factor(covariates$region)
```

```
GPS_mod <-xgboost(data = data.matrix(covariates[,c(4:19)]),
                label = covariates$pm25_ensemble,nrounds=50)
mod_sd<- sd(covariates$pm25_ensemble - predict(
        GPS_mod,data.matrix(covariates[,c(4:19)])))
feature_names <- GPS_mod$feature_names
covariates$GPS<-dnorm(covariates$pm25_ensemble, mean = predict(
        GPS_mod, data.matrix(covariates[feature_names])), sd = sd(
        covariates$pm25_ensemble - predict(
        GPS_mod, data.matrix(covariates[feature_names]))))
```

S.4.5 Code for Implementing GPS Matching Approaches

```
# Causal Modelling: Matching by GPS
# Matching algorithm on single exposure level a
matching.fun.dose.l1.caliper2 <- function(simulated.data,
                                          GPS_mod,
                                          a,
                                          delta_n=1,
                                          scale)
{
  simulated.data[["treat"]] <- simulated.data[["pm25_ensemble"]]
  simulated.data[["year_fac"]] <- as.factor(simulated.data[["year"]])
  simulated.data[["region"]] <- as.factor(simulated.data[["region"]])
  p.a <- dnorm(a,mean = predict(GPS_mod,data.matrix(
            simulated.data[feature_names])), sd=mod_sd)
  ## calculate min and max once
  treat.min <- min(simulated.data[["treat"]],na.rm=T)
  treat.max <- max(simulated.data[["treat"]],na.rm=T)
  GPS.min <- min(simulated.data[["GPS"]],na.rm=T)
  GPS.max <- max(simulated.data[["GPS"]],na.rm=T)
```

```
  ## Calculate standardized GPS and treatment
  simulated.data <- transform(simulated.data,
                    std.treat = (treat - treat.min) / (treat.max - treat.min),
                    std.GPS = (GPS - GPS.min) / (GPS.max - GPS.min))
  std.a <- (a - treat.min) / (treat.max - treat.min)
  std.p.a <- (p.a - GPS.min) / (GPS.max - GPS.min)

  simulated.data.subset <- simulated.data[abs(simulated.data[["treat"]] - a)
                <= (delta_n/2), ]
  ## Find the nearest neighbor matching
  wm <- apply(abs(outer(simulated.data.subset[["std.GPS"]], std.p.a, `-`))* scale,
              2, function(x) which.min(abs(simulated.data.subset[["std.treat"]] -
              std.a)* (1 - scale) + x)
  )
  dp <- simulated.data.subset[wm, c("dead", "time_count")]
  E.a <- apply(dp, 2, sum, na.rm = T)
  return(c(simulated.data[1,3:7], E.a[1], E.a[2], a))
}
# function to implement the matching for each strata
par_match <- function(a_i=a_i,
                      data.list,
                      GPS_mod=GPS_mod,
                      delta_n=delta_n,
                      scale=scale){
   matching_level <- data.table(Reduce(rbind,mclapply(1:length(data.list),
   function(i, a_i=a_i, GPS_mod=GPS_mod, delta_n=delta_n, scale=scale){
    return(matching.fun.dose.l1.caliper2(simulated.data=data.list[[i]],
    GPS_mod=GPS_mod, a=a_i, delta_n=delta_n, scale=scale))
    }, GPS_mod=GPS_mod, a_i=a_i, delta_n=delta_n, scale=scale, mc.cores=cl)))
   colnames(matching_level) <-c("sex", "race", "dual", "entry_age_break",
```

```
   "followup_year", "dead", "time_count", "pm25_ensemble")
  return(matching_level)
}
delta_n<-a.vals[2]-a.vals[1]
matching<-par_match(a.vals + delta_n/2, data.list=aggregate_data.list,
    GPS_mod=GPS_mod, delta_n=delta_n, scale=1)
matching_gnm <- summary(gnm(dead ~ pm25_ensemble+offset(log(time_count)),
    eliminate = (as.factor(sex):as.factor(race):as.factor(dual):
    as.factor(entry_age_break):as.factor(followup_year)),
    data=subset(matching, family=poisson(link="log")))
```

S.4.6 Code for Implementing GPS Weighting Approaches

```
# Causal Modelling: Weighting by GPS
Nm <- dnorm(covariates$pm25_ensemble, mean = mean(
        covariates$pm25_ensemble,na.rm=T), sd = sd(
        covariates$pm25_ensemble,na.rm=T))
covariates$IPW <- Nm/(covariates$GPS)
covariates <- covariates[,c("zip","year","IPW","GPS")]
aggregate_data<-merge(aggregate_data,covariates,by=c("zip","year"),all.x=T)

aggregate_data$IPW[aggregate_data$IPW>10]<-10
IPTW_gnm <- gnm(dead ~ pm25_ensemble + offset(log(time_count)),
    eliminate=(as.factor(sex):as.factor(race):as.factor(dual):
    as.factor(entry_age_break):as.factor(followup_year)),
    data=aggregate_data, family = poisson(link="log"),weights = IPW)
```

S.4.7 Code for Implementing GPS Adjustment Approaches

```
# Causal Modelling: Adjustment by GPS
a.vals <- seq(min(aggregate_data$pm25_ensemble), max(
```

```
aggregate_data$pm25_ensemble), length.out = 50)
GPScova.fun.dose <- function(simulated.data,
                             GPS_mod,
                             a,
                             model){
    simulated.data$year_fac <- as.factor(simulated.data$year)
    simulated.data$region <- as.factor(simulated.data$region)
    p.a <- dnorm(a, mean = predict(GPS_mod,data.matrix(
        simulated.data[feature_names])), sd = sd(simulated.data$pm25_ensemble -
            predict(GPS_mod,data.matrix(simulated.data[feature_names]))))
    data.a <- data.frame(cbind(cbind(pm25_ensemble=a,GPS=p.a),
            simulated.data[,c(1:7,9)]))
    data.a$dead <- predict(model, data.a, type="response")

    data.a<-aggregate(cbind(data.a$dead,data.a$time_count,
        data.a$pm25_ensemble), by=list(data.a$sex, data.a$race, data.a$dual,
        data.a$entry_age_break, data.a$followup_year), FUN=mean)
    colnames(data.a)[1:8]<-c("sex", "race", "dual", "entry_age_break",
        "followup_year", "dead","time_count","pm25_ensemble")
    return(data.a)
}
cl<-makeCluster(8)
registerDoParallel(cl)
flexible_model<-bam(dead ~ pm25_ensemble + GPS + pm25_ensemble*GPS + I(GPS^2) +
    offset(log(time_count)) + as.factor(sex) + as.factor(race) + as.factor(dual) +
    as.factor(entry_age_break)+as.factor(followup_year), data=aggregate_data,
    family=poisson(link="log"), chunk.size=5000, cluster=cl)
stopCluster(cl)
delta_n <-a.vals[2]-a.vals[1]
GPScova<-data.table(Reduce(rbind,
```

```
    lapply(a.vals+delta_n/2,GPScova.fun.dose, simulated.data=aggregate_data,
        GPS_mod=GPS_mod, model=flexible_model)))
GPScova_model<-summary(gnm(log(dead)~ pm25_ensemble + offset(log(time_count)),
    eliminate = (as.factor(sex):as.factor(race):as.factor(dual):
    as.factor(entry_age_break):as.factor(followup_year)), data=GPScova))
```

S.4.8 Code for Assessing Covariate Balance

```
# Assess covariate balance via absolute correlation
library("wCorr")
library("polycor")
## absolute correlation for unadjusted observational data
cor_origin <-
  c(abs(polyserial(covariates$pm25_ensemble,covariates$year_fac)),
    abs(polyserial(covariates$pm25_ensemble,covariates$region))),
    abs(cor(covariates$pm25_ensemble,covariates$mean_bmi,
        method = c("spearman"))),
    abs(cor(covariates$pm25_ensemble,covariates$smoke_rate,
        method = c("spearman"))),
    abs(cor(covariates$pm25_ensemble,covariates$hispanic,
        method = c("spearman"))),
    abs(cor(covariates$pm25_ensemble,covariates$pct_blk,
        method = c("spearman"))),
    abs(cor(covariates$pm25_ensemble,covariates$medhouseholdincome,
        method = c("spearman"))),
    abs(cor(covariates$pm25_ensemble,covariates$medianhousevalue,
        method = c("spearman"))),
    abs(cor(covariates$pm25_ensemble,covariates$poverty,
        method = c("spearman"))),
    abs(cor(covariates$pm25_ensemble,covariates$education,
```

```
        method = c("spearman"))),
    abs(cor(covariates$pm25_ensemble,covariates$popdensity,
        method = c("spearman"))),
    abs(cor(covariates$pm25_ensemble,covariates$pct_owner_occ,
        method = c("spearman"))),
    abs(cor(covariates$pm25_ensemble,covariates$summer_tmmx,
        method = c("spearman"))),
    abs(cor(covariates$pm25_ensemble,covariates$summer_rmax,
        method = c("spearman"))),
    abs(cor(covariates$pm25_ensemble,covariates$winter_tmmx,
        method = c("spearman"))),
    abs(cor(covariates$pm25_ensemble,covariates$winter_rmax,
        method = c("spearman"))))
## absolute correlation for weighted data
cor_weight <-
  c(abs(weightedCorr(weights=covariates$IPW2, covariates$pm25_ensemble,
        covariates$year_fac,method = c("Polyserial"))),
    abs(weightedCorr(weights=covariates$IPW, covariates$pm25_ensemble,
        covariates$region,method = c("Polyserial"))),
    abs(weightedCorr(weights=covariates$IPW2, covariates$pm25_ensemble,
        covariates$mean_bmi,method = c("spearman"))),
    abs(weightedCorr(weights=covariates$IPW2,covariates$pm25_ensemble,
        covariates$smoke_rate,method = c("spearman"))),
    abs(weightedCorr(weights=covariates$IPW2,covariates$pm25_ensemble,
        covariates$hispanic,method = c("spearman"))),
    abs(weightedCorr(weights=covariates$IPW2,covariates$pm25_ensemble,
        covariates$pct_blk,method = c("spearman"))),
    abs(weightedCorr(weights=covariates$IPW2,covariates$pm25_ensemble,
        covariates$medhouseholdincome,method = c("spearman"))),
    abs(weightedCorr(weights=covariates$IPW2,covariates$pm25_ensemble,
```

```
        covariates$medianhousevalue,method = c("spearman"))),
    abs(weightedCorr(weights=covariates$IPW2,covariates$pm25_ensemble,
        covariates$poverty,method = c("spearman"))),
    abs(weightedCorr(weights=covariates$IPW2,covariates$pm25_ensemble,
        covariates$education,method = c("spearman"))),
    abs(weightedCorr(weights=covariates$IPW2,covariates$pm25_ensemble,
        covariates$popdensity,method = c("spearman"))),
    abs(weightedCorr(weights=covariates$IPW2,covariates$pm25_ensemble,
        covariates$pct_owner_occ,method = c("spearman"))),
    abs(weightedCorr(weights=covariates$IPW2,covariates$pm25_ensemble,
        covariates$summer_tmmx,method = c("spearman"))),
    abs(weightedCorr(weights=covariates$IPW2,covariates$pm25_ensemble,
        covariates$summer_rmax,method = c("spearman"))),
    abs(weightedCorr(weights=covariates$IPW2,covariates$pm25_ensemble,
        covariates$winter_tmmx,method = c("spearman"))),
    abs(weightedCorr(weights=covariates$IPW2,covariates$pm25_ensemble,
        covariates$winter_rmax,method = c("spearman"))))
## absolute correlation for matched data
cor_matched <-
  c(abs(polyserial(match_data2$pm25_ensemble,match_data1$year_fac)),
    abs(polyserial(match_data2$pm25_ensemble,match_data1$region)),
    abs(cor(match_data2$pm25_ensemble,match_data1$mean_bmi,
        method = c("spearman"))),
    abs(cor(match_data2$pm25_ensemble,match_data1$smoke_rate,
        method = c("spearman"))),
    abs(cor(match_data2$pm25_ensemble,match_data1$hispanic,
        method = c("spearman"))),
    abs(cor(match_data2$pm25_ensemble,match_data1$pct_blk,
        method = c("spearman"))),
    abs(cor(match_data2$pm25_ensemble,match_data1$medhouseholdincome,
```

```
    method = c("spearman"))),
abs(cor(match_data2$pm25_ensemble,match_data1$medianhousevalue,
    method = c("spearman"))),
abs(cor(match_data2$pm25_ensemble,match_data1$poverty,
    method = c("spearman"))),
abs(cor(match_data2$pm25_ensemble,match_data1$education,
    method = c("spearman"))),
abs(cor(match_data2$pm25_ensemble,match_data1$popdensity,
    method = c("spearman"))),
abs(cor(match_data2$pm25_ensemble,match_data1$pct_owner_occ,
    method = c("spearman"))),
abs(cor(match_data2$pm25_ensemble,match_data2$summer_tmmx,
    method = c("spearman"))),
abs(cor(match_data2$pm25_ensemble,match_data2$summer_rmax,
    method = c("spearman"))),
abs(cor(match_data2$pm25_ensemble,match_data2$winter_tmmx,
    method = c("spearman"))),
abs(cor(match_data2$pm25_ensemble,match_data2$winter_rmax,
    method = c("spearman"))))
```

Table S1: Data Sources

	Source	Data
Exposure	Di et al. (9)	1 km^2 $PM_{2.5}$ predictions
Meteorological	Gridmet via Google Earth Engine	4 km^2 temperature and relative humidity predictions
Confounders	Census	zip code-level socioeconomic status (SES) variables
	CDC	county-level behavioral risk factor variables
Health	CMS	mortality, individual-level characteristics

Table S2: Characteristics for the Medicare Study Cohort from 2000 to 2012.

Variable	Entire Enrollees	Enrollees Exposed to $PM_{2.5} \leq 12 \mu g/m^3$
Number of individuals	56,095,877	26,408,536
Number of deaths	20,303,529	7,092,910
Total Person-years	415,551,432	160,701,132
Median years of follow-up	7.0	7.0
Individual-level characteristics		
Age at entry (years)		
% 65-74	76.4	83.0
% 75-84	18.0	12.9
% 85-94	5.0	3.8
% 95 or above	0.5	0.3
Mean (SD)	70.1 (7.1)	68.7 (6.4)
Sex		
% Female	56.1	54.3
% Male	43.9	45.7
Race		
% White	85.2	87.6
% Black	8.8	6.0
% Asian	1.7	1.5
% Hispanic	1.9	2.1
% North American Native	0.2	0.4
Medicaid eligibility		
% Eligible	11.7	10.6
Area-level risk factors characteristics		
% Ever smoker	47.5	47.7
% below poverty level	10.6	10.2
% below high school education	30.6	27.5
% of owner occupied housing	72.3	73.6
% Hispanic population	8.9	6.9
% Black population	8.6	9.0
Population density (person/mile^2)	1535.9 (4982.2)	1150.0 (3799.9)
Mean BMI (kg/m^2)	27.5 (1.1)	27.5 (1.1)
Median household income $(1000\ \$)$	47.4 (20.9)	48.7 (21.1)
Median house value $(1000\ \$)$	155.4 (134.2)	164.8 (140.5)
Meteorological Variables		
Summer temperature $(^\circ C)$	29.5 (3.7)	29.5 (4.1)
Winter temperature $(^\circ C)$	7.5 (7.1)	7.2 (7.7)
Summer relative humidity $(\%)$	88.3 (11.8)	86.5 (13.3)
Wummer relative humidity $(\%)$	86.4 (7.5)	86.8 (7.9)
$PM_{2.5}$ concentrations $(\mu g/m^3)$	10.4 (3.3)	8.7 (2.4)

Mean (SD) is presented for continuous variables.

Table S3: Analysis Results. Point estimates and 95 % confidence intervals of the Hazard Ratios (HR). These estimated HRs are obtained under four cohorts using five different statistical approaches (two traditional regression approaches and three causal inference approaches). The results of sensitivity analyses 1) excluding year 2) excluding meteorological variables are provided.

Cohort	Methods	Main analysis	Not adjust for year	Not adjust for meteorological variables
2000-2016	Matching	$1.068(1.054, 1.083)$	$1.089(1.075, 1.103)$	$1.077(1.063, 1.092)$
	Weighting	$1.076(1.065, 1.088)$	$1.144(1.134, 1.154)$	$1.087(1.076, 1.098)$
	Adjustment	$1.072(1.061, 1.082)$	$1.115(1.103, 1.128)$	$1.061(1.050, 1.072)$
	Cox	$1.066(1.058, 1.074)$	$1.172(1.164, 1.180)$	$1.058(1.050, 1.066)$
	Poisson	$1.062(1.055, 1.069)$	$1.166(1.158, 1.174)$	$1.057(1.049, 1.064)$
2000-2016	Matching	$1.261(1.233, 1.289)$	$1.318(1.287, 1.349)$	$1.251(1.222, 1.280)$
Low Level	Weighting	$1.268(1.237, 1.300)$	$1.387(1.355, 1.419)$	$1.262(1.232, 1.291)$
	Adjustment	$1.231(1.180, 1.284)$	$1.424(1.327, 1.527)$	$1.233(1.169, 1.299)$
	Cox	$1.369(1.340, 1.399)$	$1.569(1.536, 1.602)$	$1.358(1.330, 1.387)$
	Poisson	$1.347(1.320, 1.375)$	$1.541(1.510, 1.573)$	$1.343(1.316, 1.370)$
2000-2012	Matching	$1.055(1.042, 1.068)$	$1.085(1.072, 1.098)$	
	Weighting	$1.067(1.056, 1.079)$	$1.114(1.103, 1.125)$	
	Adjustment	$1.047(1.037, 1.057)$	$1.078(1.065, 1.090)$	
	Cox	$1.059(1.051, 1.067)$	$1.128(1.120, 1.136)$	
	Poisson	$1.055(1.048, 1.063)$	$1.123(1.116, 1.131)$	
2000-2012	Matching	$1.271(1.241, 1.301)$	$1.293(1.262, 1.324)$	
Low Level	Weighting	$1.298(1.254, 1.344)$	$1.383(1.343, 1.425)$	
	Adjustment	$1.233(1.176, 1.292)$	$1.385(1.291, 1.485)$	
	Cox	$1.367(1.331, 1.404)$	$1.538(1.497, 1.580)$	
	Poisson	$1.342(1.308, 1.377)$	$1.509(1.471, 1.548)$	

"Low level" is defined as Medicare enrollees exposed to $PM_{2.5} \leq 12\ \mu g/m^3$.

Table S4: The Importance Scores of Variables in the GPS models. The importance scores represent the fractional contribution of each variable to the model based on the total gains (28).

	Entire Medicare Enrollees		Exposed to $PM_{2.5} \leq 12\mu g/m^3$	
Variables	2000-2016	2000-2012	2000-2016	2000-2012
Area-level risk factors characteristics				
% Ever smoker	1.7×10^{-2}	1.7×10^{-2}	2.7×10^{-2}	2.2×10^{-2}
% below poverty level	3.9×10^{-4}	7.4×10^{-4}	1.1×10^{-4}	1.4×10^{-4}
% below high school education	1.9×10^{-2}	2.2×10^{-2}	6.1×10^{-3}	4.1×10^{-3}
% of owner occupied housing	4.7×10^{-3}	7.6×10^{-3}	3.8×10^{-3}	4.8×10^{-3}
% Hispanic population	9.5×10^{-3}	8.7×10^{-3}	4.3×10^{-3}	3.7×10^{-3}
% Black population	9.6×10^{-3}	8.9×10^{-3}	9.8×10^{-3}	1.2×10^{-2}
Population density (person/mile^2)	1.8×10^{-1}	2.1×10^{-1}	2.0×10^{-1}	2.1×10^{-1}
Mean BMI (kg/m^2)	1.7×10^{-2}	1.3×10^{-2}	1.6×10^{-2}	1.0×10^{-2}
Median household income $(1000\ \$)$	2.3×10^{-3}	2.3×10^{-3}	3.2×10^{-3}	4.3×10^{-3}
Median house value $(1000\ \$)$	9.4×10^{-3}	9.7×10^{-3}	1.1×10^{-2}	1.1×10^{-2}
Meteorological Variables				
Summer temperature $(^\circ C)$	1.1×10^{-1}	1.3×10^{-1}	1.3×10^{-1}	1.3×10^{-1}
Winter temperature $(^\circ C)$	9.2×10^{-2}	1.2×10^{-1}	6.4×10^{-2}	7.3×10^{-2}
Summer relative humidity $(\%)$	1.2×10^{-1}	1.3×10^{-1}	2.4×10^{-2}	2.6×10^{-2}
Winter relative humidity $(\%)$	2.2×10^{-2}	2.1×10^{-2}	1.8×10^{-2}	1.7×10^{-2}
Census Region	8.5×10^{-2}	1.0×10^{-1}	3.0×10^{-1}	3.8×10^{-1}
Year	3.1×10^{-1}	2.0×10^{-1}	1.7×10^{-1}	9.0×10^{-2}

Table S5: E-value for point estimates and the lower bound of the 95% confidence intervals of the Hazard Ratios (HR). These estimated HRs are obtained under four cohorts using five different statistical approaches (two traditional regression approaches and three causal inference approaches). The results of analogous E-value 1) for year 2) for meteorological variables are also provided.

Cohort	Methods	Main analysis	E-value for year	E-value for meteorological variables
2000-2016	Matching	1.34 (1.29)	1.16	1.10
	Weighting	1.36 (1.33)	1.32	1.11
	Adjustment	1.35 (1.32)	1.24	1.11
	Cox	1.33 (1.31)	1.43	1.09
	Poisson	1.32 (1.30)	1.43	1.07
2000-2016	Matching	1.83 (1.77)	1.26	1.10
Low Level	Weighting	1.85 (1.78)	1.41	1.07
	Adjustment	1.76 (1.64)	1.58	1.04
	Cox	2.08 (2.01)	1.56	1.10
	Poisson	2.03 (1.97)	1.55	1.06
2000-2012	Matching	1.30 (1.25)	1.20	
	Weighting	1.33 (1.30)	1.26	
	Adjustment	1.27 (1.23)	1.20	
	Cox	1.31 (1.28)	1.33	
	Poisson	1.30 (1.27)	1.33	
2000-2012	Matching	1.86 (1.79)	1.15	
Low Level	Weighting	1.92 (1.82)	1.33	
	Adjustment	1.77 (1.63)	1.50	
	Cox	2.08 (1.99)	1.50	
	Poisson	2.02 (1.94)	1.50	

"Low level" is defined as the analysis restricted to Medicare enrollees exposed to $PM_{2.5} \leq 12$ $\mu g/m^3$.

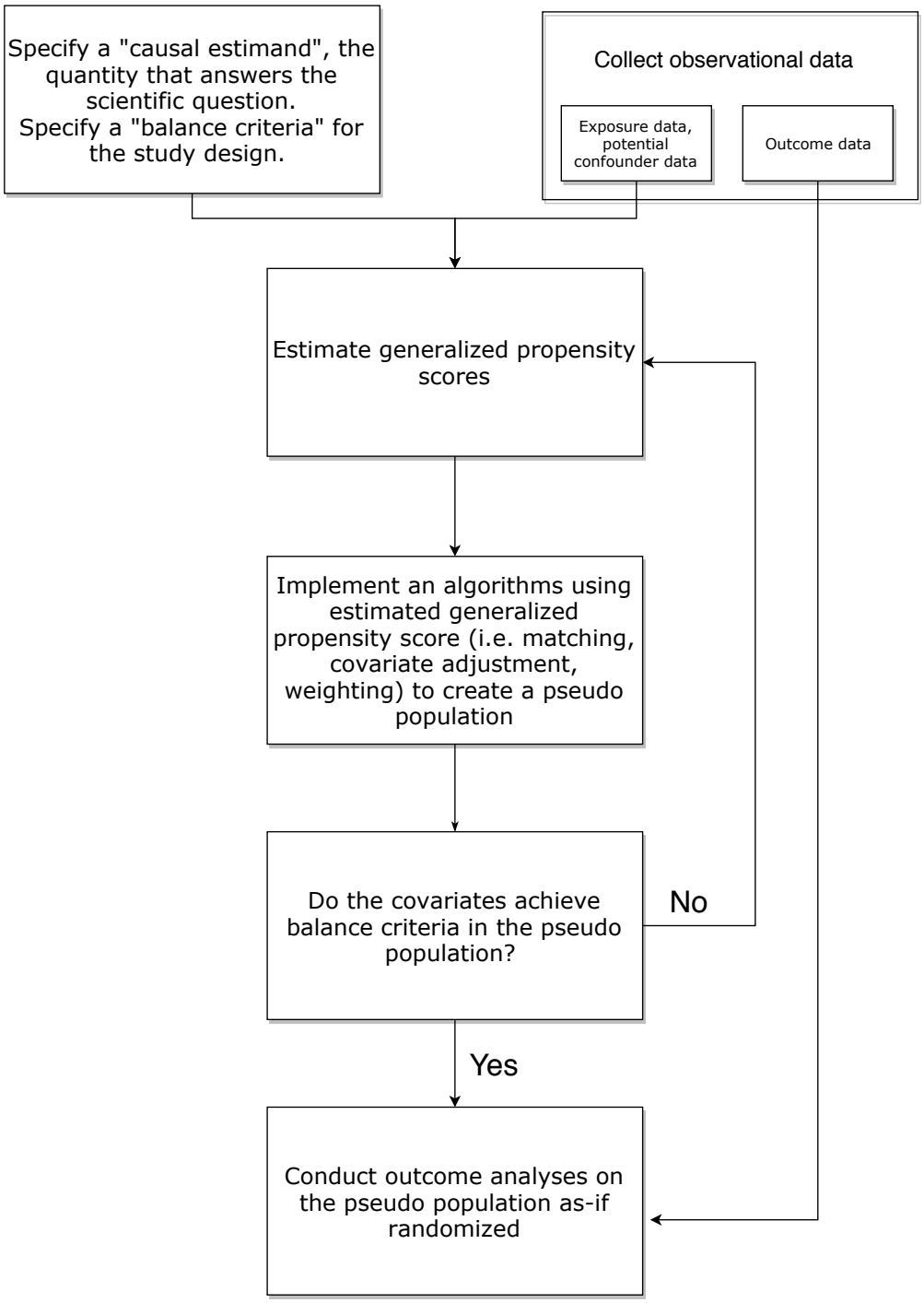

Figure S1: Causal Inference Workflow. A workflow for causal inference approaches using generalized propensity scores to design and analyze observational data. The design and analysis stages are kept separate, and the outcome information is not used to construct the pseudo-population in the design stage. The technical details of implementing each of the steps are discussed in Section **S.**1. Additional technical considerations for each step are described in detail in the causal inference literature (47; 64; 65).

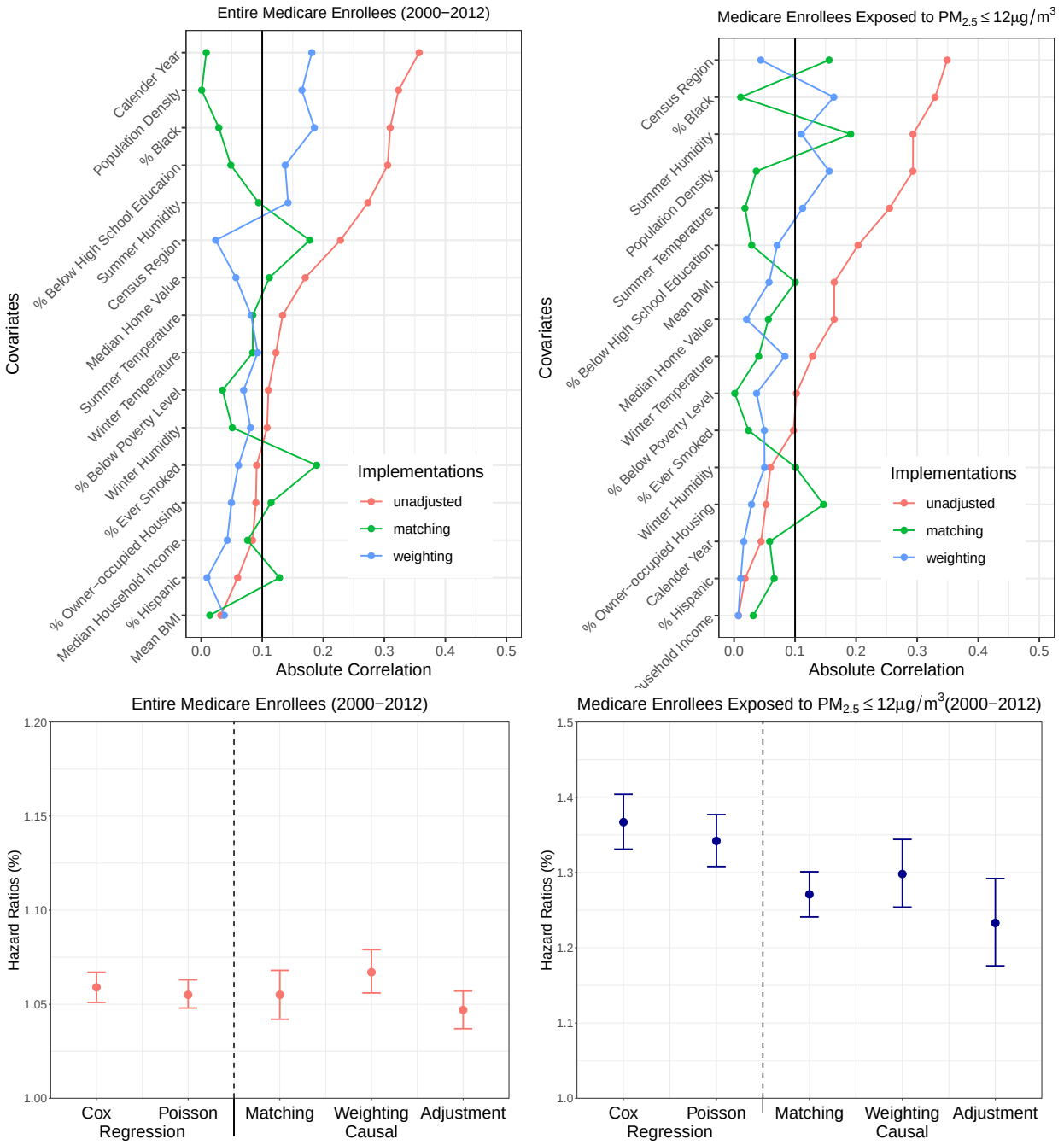

Figure S2: Absolute Correlations (ACs), Point estimates and 95% Confidence Intervals of the Hazard Ratios (HR) for the Study Cohort from 2000 to 2012. The upper panels represent the ACs for each covariate in the weighted (blue line), matched population (green line) and unadjusted observational population (red line). The black line represents the cut-off of covariate balance suggested by Zhu et al. (29). In general, weighting and matching substantially improve covariate balance for these potential confounders. The average AC is 0.17 in all Medicare enrollees, 0.08 after matching and 0.09 after weighting. The average AC is 0.16 in Medicare enrollees who were exposed to $PM_{2.5} \leq 12\ \mu g/m^3$, 0.07 after matching and 0.06 after weighting (upper panel). The lower panels represents the estimated HRs obtained under five different statistical approaches (two traditional regression approaches and three causal inference approaches). They represent the risk of all-cause mortality associated with a $10\ \mu g/m^3$ increase in $PM_{2.5}$. The left panel provides results based on all Medicare enrollees 2000–2012. The right panel provides results based on Medicare enrollees who were exposed to $PM_{2.5}$ lower than $12\ \mu g/m^3$ 2000–2012.

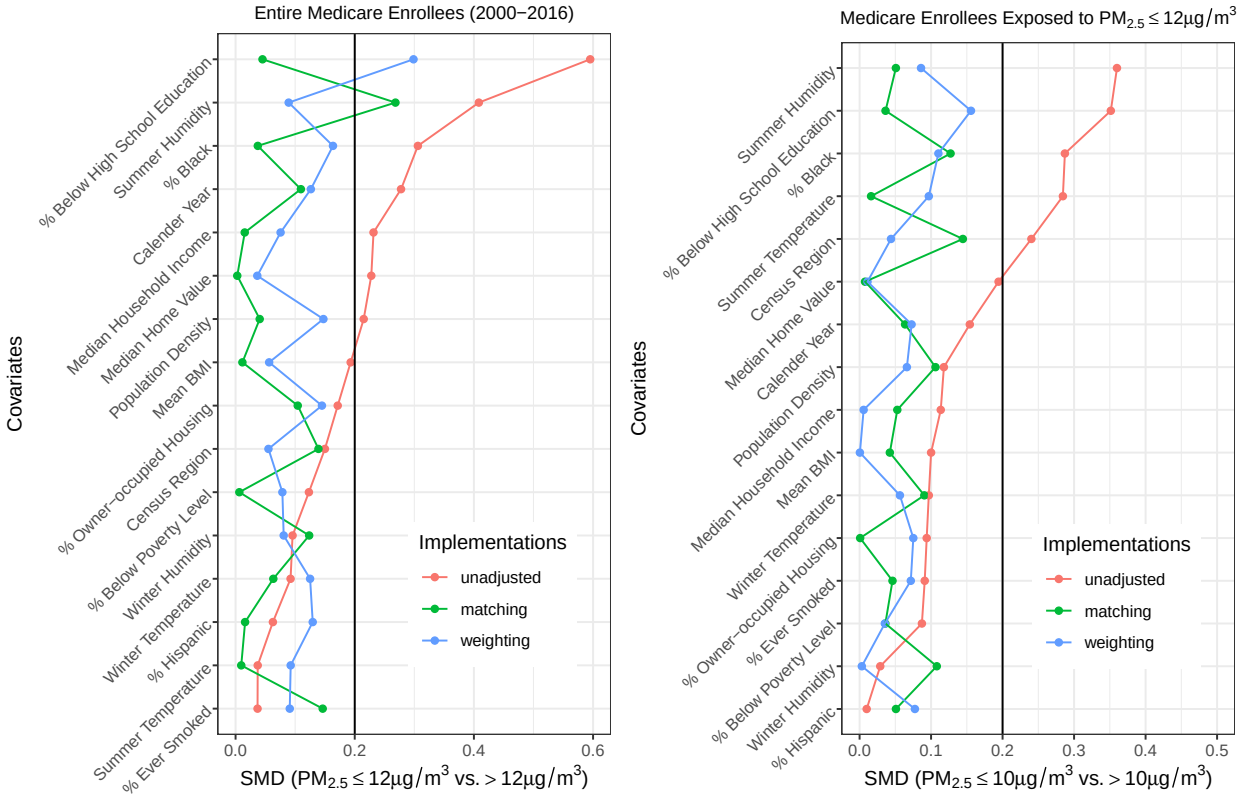

Figure S3: Standardized mean differences (SMDs) for Study Cohort from 2000 to 2016. The figure represents the SMDs for each covariate in the weighted (blue line), matched population (green line) and unadjusted observational population (red line). The black line represents the cut-off of covariate balance suggested by Harder et al.(46). In general, weighting and matching substantially improved covariate balance for these potential confounders.

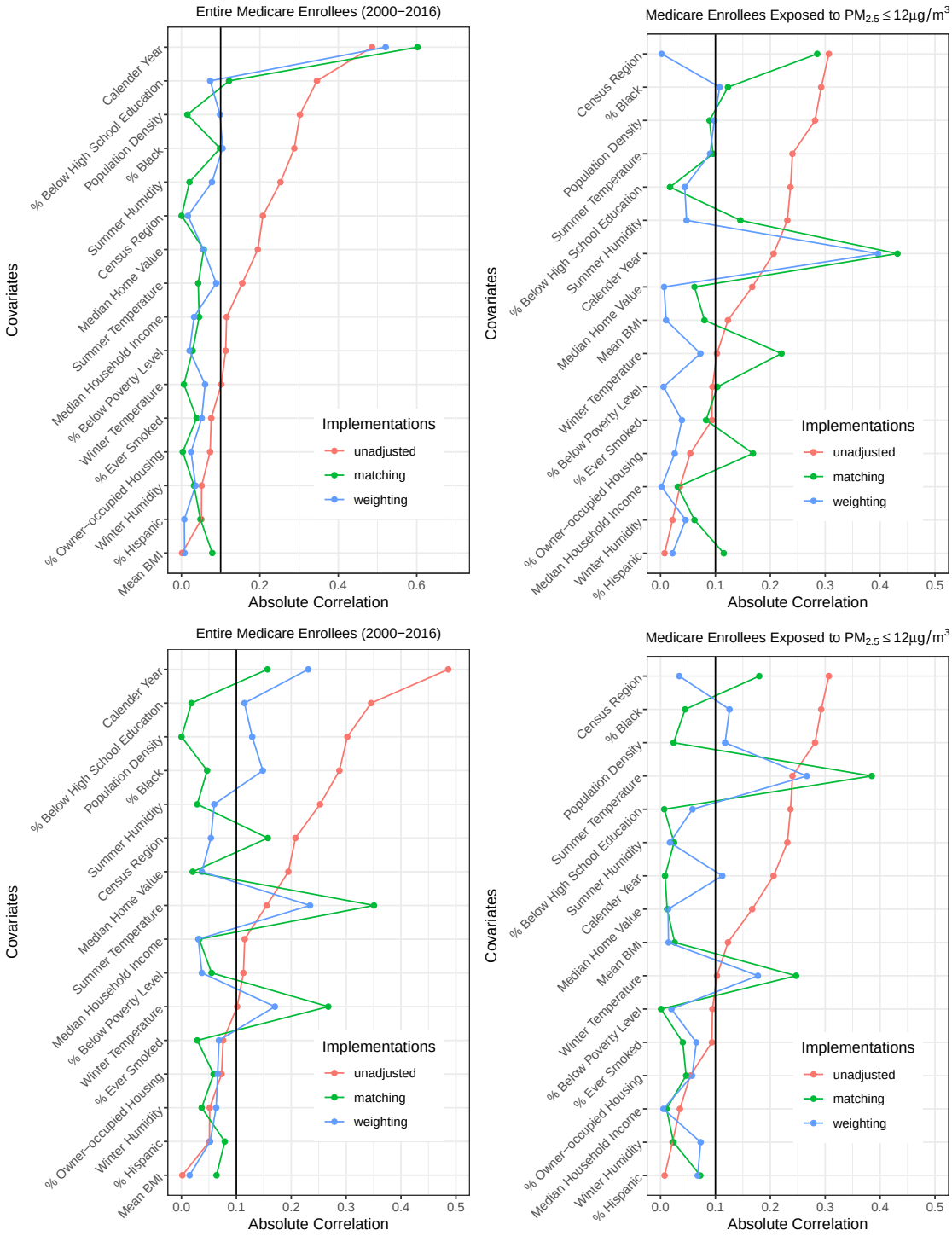

Figure S4: Absolute Correlations (ACs) for Study Cohort from 2000 to 2016 Excluding Year or Meteorological Variables as Confounders in the GPS model. The upper panels represent the ACs for each covariate in the weighted (blue line), matched population (green line) and unadjusted observational population (red line) when excluding year. In general, weighting and matching substantially improve covariate balance for potential confounders included in the GPS model, yet remain largely imbalanced for the year variable. For Medicare enrollees who were always exposed to $PM_{2.5}$ lower than 12 $\mu g/m^3$ 2000-2016, the matching approach does not perform well and four variables remain largely imbalanced. The lower panels represent the ACs for each covariate when excluding meteorological variables. Weighting and matching substantially improve covariate balance for potential confounders included in the GPS model, yet remain largely imbalanced for the meteorological variables, particularly summer and winter temperature. The black line represents the cut-off of covariate balance suggested by Zhu et al. (29).

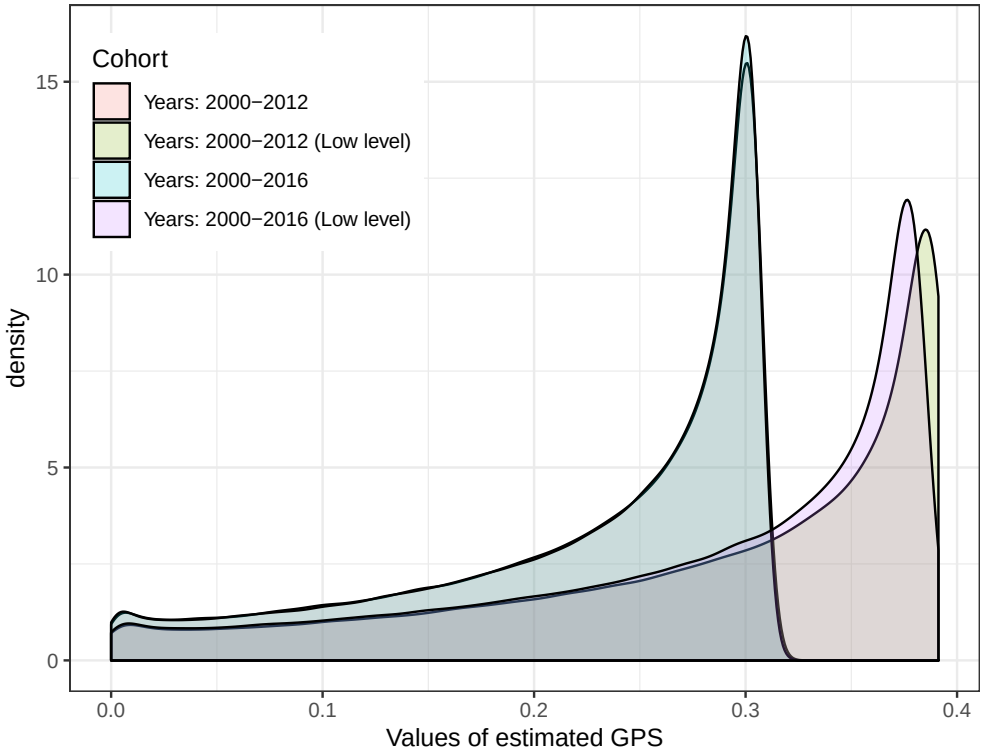

Figure S5: Estimated Values of GPS. Red represents the estimated GPS in all Medicare enrollees among years 2000-2012, green represents the estimated GPS in Medicare enrollees who were always exposed to low level $PM_{2.5}$ among years 2000-2012, blue represents the estimated GPS in all Medicare enrollees among years 2000-2016, and purple represents the estimated GPS in Medicare enrollees who were always exposed to low level $PM_{2.5}$ among years 2000-2016. The GPS estimation is conducted using machine learning approaches called gradient boosting machine (28). "Low level" is defined as Medicare enrollees exposed to $PM_{2.5} \leq 12\ \mu g/m^3$.