Environmental Health

Assessing short-term impact of PM_{10} on mortality using a semiparametric generalized propensity score approach

Laura Forastiere[1,2], Michele Carugno[3] and Michela Baccini[1*]

Abstract

Background: The shape of the exposure-response curve describing the effects of air pollution on population health has crucial regulatory implications, and it is important in assessing causal impacts of hypothetical policies of air pollution reduction.

Methods: After having reformulated the problem of assessing the short-term impact of air pollution on health within the potential outcome approach to causal inference, we developed a method based on the generalized propensity score (GPS) to estimate the average dose-response function (aDRF) and quantify attributable deaths under different counterfactual scenarios of air pollution reduction. We applied the proposed approach to assess the impact of airborne particles with a diameter less than or equal to 10 μm (PM_{10}) on deaths from natural, cardiovascular and respiratory causes in the city of Milan, Italy (2003-2006).

Results: As opposed to what is commonly assumed, the estimated aDRFs were not linear, being steeper for low-moderate values of exposure. In the case of natural mortality, the curve became flatter for higher levels; this behavior was less pronounced for cause-specific mortality. The effect was larger in days characterized by higher temperature. According to the curves, we estimated that a hypothetical intervention able to set the daily exposure levels exceeding 40 μg/m^3 to exactly 40 would have avoided 1157 deaths (90%CI: 689, 1645) in the whole study period, 312 of which for respiratory causes and 771 for cardiovascular causes. These impacts were higher than those obtained previously from regression-based methods.

Conclusion: This novel method based on the GPS allowed estimating the average dose-response function and calculating attributable deaths, without requiring strong assumptions about the shape of the relationship. Its potential as a tool for investigating effect modification by temperature and its use in other environmental epidemiology contexts deserve further investigation.

Keywords: Short-term effects of air pollution, Health impact assessment, Attributable deaths, Generalized propensity score, Potential outcomes, Exposure-response function

*Correspondence: michela.baccini@unifi.it
[1]Department of Statistics, Computer Science, Applications, University of Florence, Viale Morgagni 59, 50134 Florence, Italy
Full list of author information is available at the end of the article

Background

In investigating short-term effects and impacts of air pollution on population mortality and morbidity, the exposure-response relationship is frequently assumed to be linear on a logarithmic scale, supporting the idea that an increase of the air pollution level yields the same percent increase in the occurrence of health events at any exposure level. This choice is in line with part of the epidemiological literature which does not report evidences of strong deviations from log-linearity within the ranges of exposure levels observed in urban areas [1–10]. However, recent works suggest a possible violation of the log-linearity assumption, especially in contexts where the range of exposure values was sufficiently wide to allow exploring the curve also for high air pollutant levels. For example, in a study aimed at investigating the relationship between the concentration of atmospheric particulate matter having diameter less than or equal to 2.5 μm ($PM_{2.5}$) and mortality in 272 Chinese cities, a leveling off in the exposure-response curves at high concentrations was reported in some areas of the Country [11]. Another Chinese study found a non log-linear increase in the occurrence of respiratory diseases for daily $PM_{2.5}$ levels above 50 μg/m^3 [12]. Evidences of violation of the log-linearity assumption were found also in Li et al. [13] for the relationship between $PM_{2.5}$ and mortality and in Zu et al. [14] for the relationship between ozone and hospitalizations from asthma. Finally, in the recent study conducted by Liu and colleagues on 652 cities around the world, the estimated pooled exposure-response functions between daily average levels of PM_{10} and $PM_{2.5}$ and all-cause mortality showed no thresholds and seemed to flatten at high air pollutants concentrations [15].

The shape of the exposure-response relationship has important regulatory implications. As a matter of fact, in the absence of a safe threshold below which exposure has no effect on mortality (which is the case of a log-linear exposure-response relationship), any reduction of the air pollutant concentrations is expected to bring to a reduction of the mortality/morbidity burden attributable to the exposure, even when air pollution levels are already under the actual regulatory limits: as such, this suggests the opportunity of policies even in areas characterized by low-moderate levels of air pollution. On the other hand, should there be a safe threshold, regulations reducing concentrations already below such a threshold would be considered to have a small impact on public health [10]. This is also why the shape of the exposure-response relationship, as related to the existence of a safe threshold for air pollutant concentration, has been a matter of debate both within the scientific community and across regulatory agencies [16, 17].

A second aspect related to log-linearity concerns the quantification of mortality and disease burdens in the population. In fact, it was shown that even in case of mild violation of log-linearity, assuming log-linearity could lead to under- or over-estimate the impact of air pollution, intended as number of health events attributable to the exposure [6]. Despite this, it is still common practice to estimate impacts assuming a log-linear effect of the exposure (see for example [15]). Gasparrini and Leone [18] did propose a method to estimate the fraction of health events attributable to extreme ambient temperatures from a flexible exposure-response curve. However, their approach, although potentially applicable to assess also the short term impact of air pollution, is not yet common practice in this field.

Up to now, the studies that explored the violation of the log-linearity assumption in the context of the short-term effects of air pollution usually described the exposure-response function by including flexible terms in the Poisson or quasi-Poisson regression model specified on the daily number of health events, for example a regression spline or a penalized spline defined on the daily air pollutant concentrations [7, 12, 14, 15, 19, 20]. Alternative approaches to regression have never been adopted.

The potential outcomes approach to causal inference [21, 22] was recently considered a useful tool in the policy debate about air pollution regulatory interventions [23], as it encourages to think in terms of causes and effects, within a formal mathematical framework. The causal interpretation of the estimated associations is particularly relevant when the effect estimates are used to calculate the absolute mortality and disease burden related to the air pollution exposure. Some studies adopted causal inference methods in the analysis of short-term effects of air pollution on population health, but none of these explored the shape of the exposure-response function [24, 25].

The generalized propensity score (GPS) method allows a flexible modeling of the exposure-response function within a potential outcomes approach to causal inference. The GPS was introduced by Hirano and Imbens [26] as a generalization of the propensity score (PS) – used in the case of a binary treatment – to the case of discrete treatments, continuous treatments and arbitrary treatment regimes [27–30]. As opposed to standard regression adjustment, the GPS-based adjustment allows adjusting for the potentially non-linear effect of a large number of covariates and it results in a flexible way to explore effect modification with every characteristic.

In the present paper, we proposed to use a semiparametric GPS method for covariate adjustment to estimate the exposure-response function describing the short-term effect of the air pollutant exposure on the health outcome, in the context of epidemiological time series analyses. In particular, we built on the semiparametric GPS proposed by Bia and colleagues [31], replacing the normal model for the outcome with a quasi-Poisson model, more suitable

for count data, and including spline terms to account for non-linearities. In addition, we implemented a novel procedure for the estimation of the attributable number of events according to the estimated exposure-response curve, assuming different counterfactual scenarios.

In this paper, we also defined ad hoc estimands to investigate the role of temperature in modifying the effect of air pollution. As a matter of facts, a major point in the study of the short-term effects of air pollution on health is the possible synergic effect between exposure to air pollution and ambient temperature. This issue is particularly relevant if we consider that, according to the climate change projections, number and intensity of extreme weather days are expected to increase in the future [32]. In the literature, there is some evidence that the association between air pollution and mortality is stronger during summer or in days with high temperatures [33–36], but further investigation is needed.

We applied the proposed approach to the estimation of short-term effects and impacts of particulate matter with a diameter less than or equal to 10 μm (PM_{10}) on mortality from all natural, respiratory, and cardiovascular causes in the city of Milan, Italy for the period 2003-2006. This data set has already been analyzed according to the standard method based on Poisson regression [37, 38] and, more recently, using propensity score-based matching with a dichotomous version of the exposure (high levels versus low levels of PM_{10}) [25]. This allowed us to compare our results to the findings of these previous studies.

Methods
Data
We considered data for the city of Milan (1,299,633 inhabitants in 2007) for the years 2003-2006. Milan is located in an area where unfavorable geographical and climate conditions induce frequent phenomena of thermal inversion, exacerbating the air pollution mainly due to traffic and, to a minor extent, to non-industrial combustion [39]. The daily time series of PM_{10} levels, temperature and humidity were obtained by averaging data over the available air quality monitoring stations of the Regional Environmental Protection Agency (ARPA Lombardia). Daily mortality data were obtained from the Regional Mortality Register. We focused on the deaths of the resident population occurred within the city area. We considered daily mortality from all natural (International Classification of Diseases, Ninth Revision (ICD-9) codes below 800), cardiovascular (ICD-9: 390-459) and respiratory (ICD-9: 460-519) causes.

Notation and potential outcomes
Let $i = 1, \ldots, N$ be the indicator of the day, which will be also referred to as the unit. Let $Z_i \in \mathcal{Z}$ be the exposure level in day i, defined as the average level of PM_{10} in

the current day i and in the previous one $i - 1$ (lag 0-1 exposure), and let $Y_i \in \mathcal{Y}$ be the number of deaths in day i. Finally, let $\mathbf{X}_i \in \mathcal{X}$ be a vector of K covariates for day i, which includes meteorological variables (ambient temperature and humidity), seasonality terms, holidays and influenza epidemics indicators.

According to the potential outcome framework [21, 22], we denote by $Y_i(z)$ the potential number of deaths in day i if z were the exposure level in that day. For each day a collection of potential outcomes is defined, one for each possible level of exposure z, but we only observe the one corresponding to the actual exposure of that day, Z_i, being $Y_i(Z_i) = Y_i$. Thus, our effort is to define a procedure that allows us to extrapolate information on the unobserved potential outcomes across days with similar baseline covariates.

Note that potential outcomes of the form $Y_i(z)$, with $z \in \mathcal{Z}$ are well-defined only under the Stable Unit Treatment Value Assumption (SUTVA) [22], which rules out the presence of interference between units, i.e., the possibility that the potential outcomes of one day are affected by the level of exposure in previous days. In this paper, we considered the lag 0-1 exposure not only to allow comparison with previous results, but also to prevent interference across days as required by SUTVA. In fact, using the lag 0-1 exposure instead of the current PM_{10} level makes the no-interference assumption more plausible [25]. Enlarging the window of the moving average for the exposure definition would have made the no-interference assumption even more plausible, but at the price of a lower variability of the exposures and of a reduced possibility of detecting an impact.

Causal estimands
In our analysis we focused on different causal estimands. We defined the average dose-response function (aDRF) as:

$$\mu(z) = \frac{1}{N} \sum_i Y_i(z), \tag{1}$$

which represents the average daily number of deaths we would have observed fixing the exposure to z in each day during the study period. We estimated this quantity precisely to capture the average causal relationship between the exposure level and mortality rate in the city of Milan during the years 2003-2006. We also defined conditional average dose-response functions of the type:

$$\mu(z; \mathbf{x}) = \frac{1}{N(\mathcal{X}^\star)} \sum_{i: \mathbf{X}_i \in \mathcal{X}^\star} Y_i(z), \tag{2}$$

where $N(\mathcal{X}^\star)$ is the number of days with covariate levels included in the subset $\mathcal{X}^\star$ of $\mathcal{X}$. We used this kind of estimand to investigate how the relationship between

exposure and outcome changed according to the levels of air temperature.

In addition to the average dose-response functions, we also focused on the absolute number of deaths attributable to observed levels of exposure exceeding a pre-specified counterfactual threshold $z^\star$. In the epidemiological literature, this quantity is usually referred to as *attributable deaths (AD)* and can be expressed as:

$$AD(z^\star) = \sum_{i:Z_i>z^\star} (Y_i - Y_i(z^\star)). \qquad (3)$$

In other words, $AD(z^\star)$ compares the observed mortality in days with $Z_i > z^\star$ to the mortality that we would have observed had the exposure been set exactly equal to $z^\star$.

As an alternative to $AD(z^\star)$, we quantified the number of attributable deaths with reference to a more complex, but realistic, counterfactual scenario, reflecting a hypothetical intervention that, instead of fixing the exposure to a specific value $z^\star$, would replace the observed exposures when $Z_i > z^\star$ with exposure values drawn from an observed or hypothetical distribution $p^\star(z)$. This new causal estimand, referred to as *distributional attributable deaths (DAD)*, takes the following form:

$$DAD(p^\star(z), z^\star) = \int_z \left(\sum_{i:Z_i>z^\star} (Y_i - Y_i(z)) \right) p^\star(z) dz. \qquad (4)$$

In our analysis, we defined $p^\star(z)$ as the actual distribution of the exposure in the days where $Z_i \leq z^\star$, that is, $p^\star(z) = p(z|Z_i \leq z^\star)$. For simplicity, we denoted the corresponding estimand by $DAD(z^\star)$.

Under the assumption that all relevant confounders have been measured, also known as the unconfoundedness assumption or the no-unmeasured confounders assumption[1], the aDRF and the other estimands presented in the previous section were identifiable and could be estimated without bias conditioning on covariates. In our analysis, the validity of the unconfoundedness assumption, which cannot be tested on data, relied on the fact that the selection of the covariates $\mathbf{X}_i$ was based on a substantive a priori knowledge of the phenomenon derived from the literature on the analysis of short term effects of air pollution, from the seminal papers [40–42] to the very recent work by Liu et al. [15].

Generalized propensity score
According to Hirano and Imbens [26], the GPS, denoted by $r(z; \mathbf{x})$, is defined as the conditional density of the exposure Z given the covariates $\mathbf{X}$:

$$r(z; \mathbf{x}) = p_{Z|\mathbf{X}}(z|\mathbf{x}). \qquad (5)$$

Similarly to the PS, the GPS is a balancing score, meaning that within strata of units with the same value of $r(z; \mathbf{x})$, the probability that Z_i is equal to z does not depend on the covariates value [43]. The unconfoundedness assumption, coupled with the balancing property of the GPS, implies that assignment to treatment is unconfounded given the GPS[2]. In practice, this means that any bias given by differences in the distribution of covariates across groups with different exposure levels can be removed by adjusting for the GPS [26].

In this paper, we used the GPS to obtain estimates of the causal estimands described in the previous section. For this purpose, first we parametrically modeled and estimated the GPS through a model for the exposure, then we specified and estimated a model for the outcome given the GPS and the exposure, and we used it to obtain estimates of the causal quantities of interest.

Model for the exposure
We assumed a parametric log-normal model for the exposure Z_i:

$$\log(Z_i) \sim N\left(\alpha_0 + \vec{\alpha}_X^T \mathbf{X}_i, \sigma_Z\right), \qquad (6)$$

where α_0 and $\vec{\alpha}_X$ were unknown regression coefficients and σ_Z was the unknown variance. In particular, we included in the model the following regressors: indicators of day of week, holiday and influenza epidemics, a linear regression spline with 5 degrees of freedom per year and equally spaced knots on the calendar day to account for seasonality and long-term trend, a cubic regression spline for temperature at lag 0-3 with 5 degrees of freedom and knots at the quantiles, linear and quadratic terms for relative humidity, and an indicator of the July-August period to account for the reduction of the population present in the city during the summer season [15, 37].

According to model 6, the GPS for day i at the level of exposure z was thus defined as:

$$r(z; \mathbf{X}_i) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} \exp -\frac{\left(\log(z) - \alpha_0 - \vec{\alpha}_X^T \mathbf{X}_i\right)^2}{2\sigma_Z^2}. \qquad (7)$$

In the same way, we defined the actual GPS for day i, i.e. the GPS evaluated at the level of exposure actually observed, as:

$$R_i = r(Z_i; \mathbf{X}_i) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} \exp -\frac{\left(\log(Z_i) - \alpha_0 - \vec{\alpha}_X^T \mathbf{X}_i\right)^2}{2\sigma_Z^2}. \qquad (8)$$

It is worth noticing that different specifications were possible for the model 6 and some effort was needed

---

[1] The weak unconfoundedness assumption states the conditional independence between exposure and potential outcomes:
$Y_i(z) \perp\!\!\!\perp Z_i \mid \mathbf{X}_i \quad \forall z \in \mathcal{Z}, \forall i.$

[2] Unconfoundedness given the GPS: $Y_i(z) \perp\!\!\!\perp Z_i \mid r(z; \mathbf{X}_i) \quad \forall z \in \mathcal{Z}, \forall i$

to find the most appropriate one. A key criterion driving the specification of the model consists in checking the balancing property of the estimated GPS. In order to check the balancing property of the GPS, we applied the method described in Hirano and Imbens [26]. After creating four classes of exposure, $C_1 = (0, 20)$, $C_2 = [20, 40)$, $C_3 = [40, 70)$, $C_4 = [70+)$, we applied the following procedure for each covariate X_i. First, we checked the independence of each covariate and each exposure class indicator $1\{Z_i \in C_k\}$ by calculating the marginal t statistics for the mean difference X_i between the class C_k and all the others as a whole. Then, we calculated for each day the GPS at the median of each treatment class (M_k, $k = 1, \ldots, 4$) and we checked the independence of each covariate X_i and each exposure class indicator $1\{Z_i \in C_k\}$, conditional on $r(M_k; X_i)$. In order to do this, given the class C_k, we defined four blocks according to the quartiles of $r(M_k, X_i)$ and, for each block, we calculated the mean difference of X_i between days belonging to the class C_k and days belonging to the other three classes, as well as the corresponding standard error. The four differences obtained, one for each GPS block, were then combined in a weighted mean, with weights proportional to the number of observations in each block. A t statistics was thus calculated as the ratio of the combined difference and its standard error. From the descriptive comparison between these GPS-adjusted t statistics and marginal t statistics (four comparisons for each variable, one for each class C_k), we were able to evaluate if adjusting for GPS improved the balance.

Model for the outcome
We specified a quasi-Poisson model for the outcome given the exposure and the GPS:

$$Y_i(z) \sim \text{quasi-Poisson}(\lambda_i) \quad \log(\lambda_i) = s\left(z, r(z; X_i)\right),$$

(9)

where $s\left(z, r(z; X_i)\right)$ was a penalized bivariate splines, with radial basis functions defined on a large number of knots[3]. The smoothing parameter was automatically selected by Generalized Cross Validation [44].

Estimation procedure
The estimation procedure consisted in the following steps:

1. We estimated the parameters of the exposure model, $\hat{\alpha}_0$, $\hat{\vec{\alpha}}_X$ and $\hat{\sigma}_Z^2$;

2. We used them to predict for each day i the actual GPS, $\widehat{R}_i$[4];
3. We used the observed data (Y_i, Z_i) and $\widehat{R}_i$ to estimate the outcome model;
4. For each exposure level z and for each unit i, we predicted the potential outcome $Y_i(z)$ using the following steps:

   a. We predicted the GPS at that level of exposure z, $\hat{r}(z; X_i)$[5];
   b. We used the estimated outcome model and $\hat{r}(z; X_i)$ to predict the potential outcome at z, $\widehat{Y}_i(z)$.

5. We used the predicted potential outcomes to estimate the quantities of interest. In particular, we estimated the average and the conditional dose-response functions by averaging the potential outcomes over all units

$$\widehat{\mu}(z) = \frac{1}{N} \sum_{i=1}^{N} \widehat{Y}_i(z)$$

and over all units with X_i included in each subset $\mathcal{X}^\star$ of interest

$$\widehat{\mu}(z, \mathbf{x}) = \frac{1}{N(\mathcal{X}^\star)} \sum_{i: X_i \in \mathcal{X}^\star} \widehat{Y}_i(z)$$

While $AD(z^\star)$ was estimated by the following sum over all units with $Z_i > z^\star$

$$\widehat{AD}(z^\star) = \sum_{i: Z_i > z^\star} Y_i - \widehat{Y}_i(z^\star),$$

a more complex procedure was needed for $DAD(z^\star)$. For the estimation of $DAD(z^\star)$ we repeated the following steps M times:

   a. We drew a value z^m from the distribution $p^\star(z)$
   b. We computed

$$\widehat{AD}^m = \sum_{i: Z_i > z^\star} \left(Y_i - \widehat{Y}_i\left(z^m\right)\right)$$

and we finally calculated the average $\widehat{DAD}(z^\star) = \frac{1}{M} \sum_m^M \widehat{AD}^m$.

Standard errors and 90% confidence intervals of the causal estimands were estimated with a bootstrap method [26, 45]. Bootstrap samples were obtained by using an independent resampling strategy with replacement. For each bootstrap sample the parameters of the exposure

---

[3]It is worth noticing that, for a fixed value of the GPS, the regression function of Y given the exposure "does not have a causal interpretation" [26]. This means that the regression model does not directly result in the causal aDRF, which instead must be derived using the procedure described in the "Estimation procedure" section

[4]$\widehat{R}_i = \frac{1}{\sqrt{2\pi\hat{\sigma}_Z^2}} \exp - \frac{\left(\log(Z_i) - \hat{\alpha}_0 - \hat{\vec{\alpha}}_X^T X_i\right)^2}{2\hat{\sigma}_Z^2}$

[5]$\hat{r}(z; X_i) = \frac{1}{\sqrt{2\pi\hat{\sigma}_Z^2}} \exp - \frac{\left(\log(z) - \hat{\alpha}_0 - \hat{\vec{\alpha}}_X^T X_i\right)^2}{2\hat{\sigma}_Z^2}$

and outcome models were estimated and causal estimands were computed using the above estimation procedure.

Results

Summary statistics of daily counts of deaths, PM_{10} concentrations and temperature levels are reported in Table 1.

In order to check the balancing property of the GPS we focused on the following variables: average daily temperature at lag 0-3, extreme temperature indicator (days with average temperature at lag 0-3 exceeding the 95th percentile calculated over the study period), humidity at lag 0, and indicators for season (summer/winter), weekends, holidays and influenza epidemics. We found that, after the adjustment for the GPS calculated according to the model specification described in the *Model for the exposure* section, the t statistics comparing the average of each covariate between classes of exposure reduced in respect to the marginal t statistics, in particular when the initial imbalance between groups was strong (Table 2). This was indicative that the balancing property was satisfied. More complex models for GPS, involving splines with a larger number of knots, a spline for humidity and interactions terms, did not bring to further improvement of the balance (results not reported).

In Fig. 1 we report the estimated aDRF for natural-cause mortality, which describes how the average level of mortality, defined as in equation 1, changes according to the PM_{10} level at lag 0-1. The relationship was increasing and nonlinear, steeper at low concentrations and flat above 50-60 μg/m^3. The aDRFs estimated for respiratory and cardiovascular mortality had a similar shape (Figs. 2 and 3). However, the tendency of these curves to flatten was less pronounced and, when present, arose at even higher PM_{10} concentrations.

Especially for natural-cause and cardiovascular mortality, the shape of the relationship was unexpectedly wiggly at low levels of exposure. This was probably due to an excessive local dependence of the curve on the knots location of the bivariate spline in equation 9, related to the fact that information under 20 μg/m^3 was quite poor, relying on only 6% of days.

For natural-cause mortality, we also estimated the aDRF within three different categories of days, defined according to daily average temperature levels (at lag 0-3): low

(up to 10°C), medium (10/23°C), high (over 23°C) (Fig. 4). As expected, background mortality was higher in days characterized by lower temperatures (blue dots) and lower in days with higher temperatures (red dots); however, the slope of the exposure-response association appeared to be slightly steeper in days belonging to the higher temperature category, suggesting a possible synergic effect between PM_{10} exposure and heat.

After estimating the aDRFs, we estimated the causal estimands introduced in the *Causal estimands* section. For the three outcomes of interest, we estimated $AD(20)$, $AD(40)$, $DAD(40)$ and $DAD(50)$ (Table 3). The choice of the threshold $z^\star$ of the countefactual scenarios followed internationally recognized standards, in details:

- 20 μg/m^3 is the World Health Orgaization (WHO) Air Quality Guideline threshold [46] for PM_{10} annual average;
- 40 μg/m^3 is the European Union (EU) limit for PM_{10} annual average [47];
- 50 μg/m^3 is the EU limit for PM_{10} daily average, not to be exceeded for more than 35 days per year [47].

Exceeding the threshold of 20 μg/m^3 has been responsible of the largest impact. In particular, an hypothetical intervention able to set the air pollution level in all days with exposure above 20 μg/m^3 to exactly 20 μg/m^3 ($AD(20)$), would have avoided 2537 deaths (90% Confidence Interval (CI): 1273, 3655), 577 (90% CI: 226, 900) of which from respiratory causes and 1000 (90% CI: 464, 1488) from cardiovascular diseases. The lowest impact was estimated for an intervention able to set PM_{10} concentration levels higher than 40 μg/m^3 to exactly 40 μg/m^3 ($AD(40)$): 1157 (90% CI: 689, 1645) attributable deaths, 312 (90% CI: 210, 418) for respiratory causes and 771 (90% CI: 580, 961) for cardiovascular causes. The fact that attributable deaths for cardiovascular and respiratory causes represented a variable proportion of the total impact was directly related to the shape of the exposure-response functions.

With $DAD(p_{40}^\star)$, we quantified the impact of a hypothetical intervention able to set the air pollution levels in all days with exposure above 40 μg/m^3 to values belonging to the PM_{10} distribution observed in days with exposure below 40 μg/m^3. As expected, $DAD(40)$ was larger

Table 1 Summary statistics for daily counts of deaths, PM_{10} concentrations and temperatures, Milan (2003-2006)

Variable	Mean	Std. Dev.	Median	Min	90th pct	Max
Natural Deaths	28.0	7.0	28	9	37	69
Cardiovascular Deaths	10.3	3.9	10	0	15	27
Respiratory Deaths	2.5	1.7	2	0	5	11
PM_{10} (μg/m^3)	52.5	32.9	43.5	3.5	99.5	227.1
Temperature (°C)	14.5	8.4	14.3	-2.0	26.0	31.8

Forastiere *et al. Environmental Health* (2020) 19:46

Page 7 of 13

Table 2 Results of the balancing check for the generalized propensity score (GPS)

Covariate	Class of exposure	Marginal *t*	GPS-adjusted *t*
Temperature (lag 0-3)	(0, 20)	5.60	-2.57
	[20, 40)	14.15	1.68
	[40, 70)	-0.39	-0.26
	[70+)	-20.78	-3.00
Influenza epidemics	(0, 20)	-2.80	-0.44
	[20, 40)	-7.47	-2.87
	[40, 70)	1.32	3.43
	[70+)	8.92	0.47
Summer indicator	(0, 20)	7.93	-2.19
	[20, 40)	14.30	2.39
	[40, 70)	-3.09	-0.77
	[70+)	-18.46	-2.48
Holidays indicator	(0, 20)	1.91	-0.07
	[20, 40)	0.02	-0.05
	[40, 70)	-2.74	-2.81
	[70+)	1.97	2.58
Humidity (lag 0)	(0, 20)	-5.78	-0.13
	[20, 40)	-5.91	1.05
	[40, 70)	-0.15	-2.78
	[70+)	10.69	1.78
Weekend indicator	(0, 20)	4.92	2.27
	[20, 40)	1.79	-1.33
	[40, 70)	-1.39	-1.13
	[70+)	-3.35	0.64
Heat episodes	(0, 20)	-2.25	-0.96
	[20, 40)	1.77	-0.15
	[40, 70)	3.48	2.43
	[70+)	-4.73	-1.99

Marginal and GPS-adjusted *t* statistics for the mean difference between one class of exposure and the others as a whole, for a set of selected covariates

than $AD(40)$, because it also picked counterfactual levels of the exposure far below the chosen threshold of 40 $\mu g/m^3$. For the same reason, also $DAD(50)$, with almost 1500 attributable deaths, returned an impact larger than $AD(40)$, despite the chosen threshold was higher. In both DAD scenarios, attributable cardiovascular deaths were about 55% of all natural attributable deaths, while respiratory mortality represented about 22% of the total.

Discussion

In this paper, we investigated the shape of the exposure-response function describing the short-term effect of PM_{10} on mortality, by using a semi-parametric procedure based on the GPS [31]. To the best of our knowledge, this is the first time that the GPS approach is applied in the analysis of short-term effects of air pollution. A recent work by Wu and colleagues [48] did propose a somewhat similar approach, but applied it to the analysis of long-term effects of air pollution; in addition, they did not quantify the mortality burden of air pollution in terms of attributable health events. On the contrary, we developed a novel procedure for the estimation of the attributable deaths based on the estimated curve, considering alternative counterfactual scenarios.

As opposed to regression-based methods, including those based on splines (e.g., Gasparrini and Leone [18]), the GPS-based method we propose allows adjusting for all possible covariates including non-linearities and interactions with the exposure - the same being more difficult in a regression model - and is less prone to incur into issues of overfitting, multicollinearity and variance inflation.

Attributable events calculation within a causal inference framework is not new in the analysis of the short-term effects of air pollution on health. In order to estimate the attributable deaths, Baccini et al. [25] implemented a PS
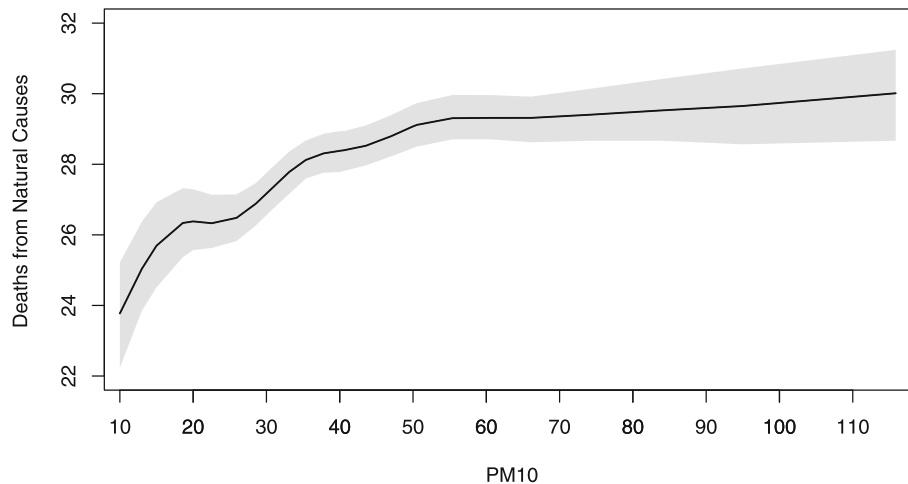

Fig. 1 Average Dose-Response Function for natural mortality. Average dose-response function (90% pointwise confidence band) of the causal relationship between PM_{10} exposure at lag 0-1 and average daily mortality from natural causes
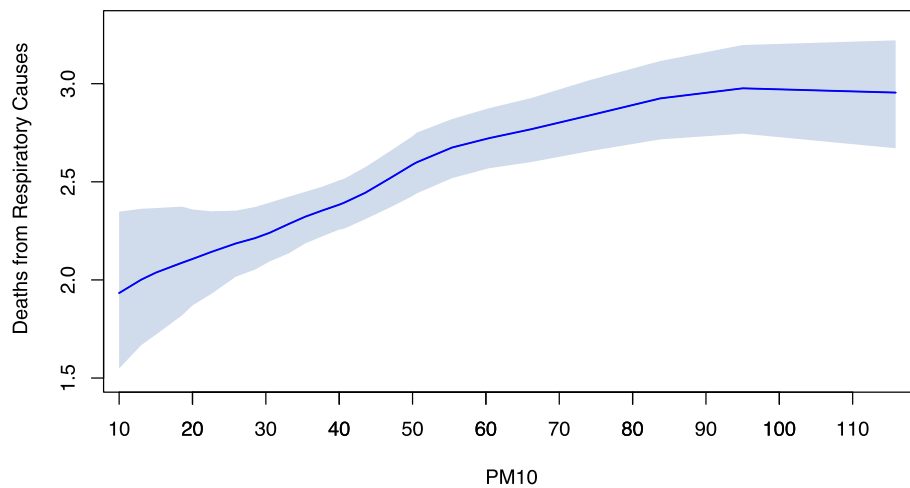
Fig. 2 Average Dose-Response Function for respiratory mortality. Average Dose-Response Function (90% pointwise confidence band) of the causal relationship between PM_{10} exposure at lag 0-1 and average daily mortality from respiratory causes

matching on the same data set used in the present paper, after dichotomizing the exposure according to a threshold of 40 $\mu g/m^3$. In that work, the level of mortality observed in days with exposure values exceeding the threshold was contrasted to the level of mortality we would have observed in those days had the levels of exposure been set to a value below the threshold. Differently, the approach proposed in the present paper does not require the definition of a binary version of the exposure, but allows to treat the air pollutant concentration as a continuous variable. Moreover, the number of attributable deaths is estimated following the shape of the (potentially nonlinear) average exposure-response curve.

In order to quantify the number of attributable events we proposed two estimands, *AD* and *DAD*. Both of them can be used to measure the impact on mortality of hypothetical interventions able to reduce the air pollutant levels (assuming that the effects of these hypothetical interventions occur only through the reduction of the air pollutant levels). Fixing a threshold for the daily exposure, *AD* compares the actual level of mortality with the level of mortality we would observe if the daily exposures exceeding the threshold were exactly equal to the threshold. On the contrary, *DAD* compares the actual level of mortality with the level of mortality we would observe if we were able to intervene on the days with daily exposures exceeding the threshold and set instead the level exposure to a value below the threshold, according to a hypothetical distribution – for example, the observed distribution of the exposures lower than the threshold.
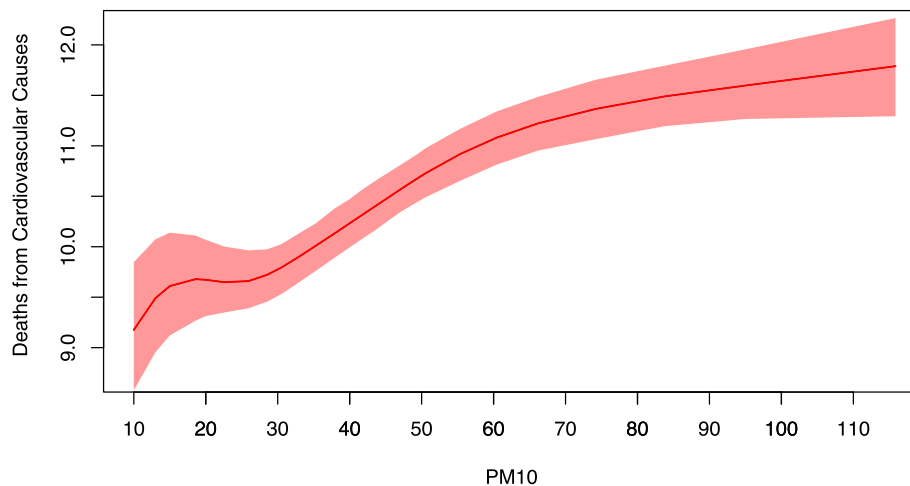

Fig. 3 Average Dose-Response Function for cardiovascular mortality. Average Dose-Response Function (90% pointwise confidence band) of the causal relationship between PM_{10} exposure at lag 0-1 and average daily mortality from cardiovascular causes
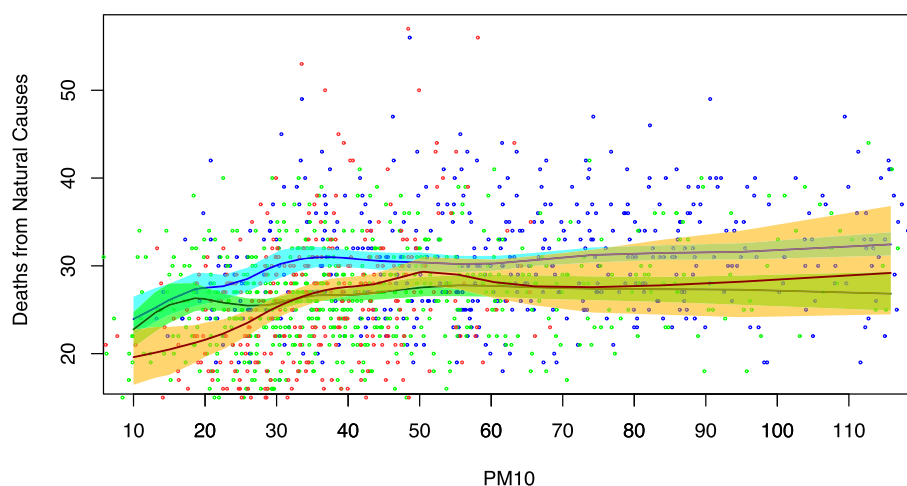
Fig. 4 Average Dose-Response Function for natural mortality by temperature level. Scatterplot and average Dose-Response Functions (90% pointwise confidence bands) of the causal relationship between PM_{10} exposure at lag 0-1 and average daily mortality from natural causes, by level of temperature at lag 0-3: up to 10°C (blue), 10/23°C (green), over 23°C (red)

It is worth noting that, while the aDRF averages the potential outcomes over all days in the study period, $AD(z^\star)$ and $DAD(z^\star)$ are sums defined only over the "treated" days, i.e. the days characterized by exposure levels higher than the threshold $z^\star$. In this sense, these quantities are somewhat related to the so-called average causal effect on the treated (ATT) [22].

Our analysis found that the estimated aDRFs were not linear. This was true also after a log transformation (not reported), thus indicating that the usual log-linearity assumption was, in this case, not valid. Moreover, being the estimated aDRFs steeper for low values of the exposure, our results seem to indicate that a "safe" threshold does not exist. As a consequence, any reduction of the PM_{10} levels, even when they are already under the actual regulatory limits, is expected to lead to a decrease in the health burden attributable to the exposure. The idea that there is no threshold, acknowledged also by governmental and international agencies [17, 46, 49], is supported by the

observation that the population is a mixture of biologically different individuals with various levels of susceptibility, and, as a consequence, no threshold below which nobody experiences the health effect can be detected, especially in the presence of frail (and potentially hypersusceptible) subjects (e.g. children, elderly people, etc.). Further considering the shape of the aDRF, while the presence of a plateau for high concentration values (as for natural-cause mortality) would suggest that small reductions of PM_{10} are ineffective in highly polluted days, the absence of a clear plateau (as for respiratory and cardiovascular mortality) suggests that any measure aimed to produce a reduction (even small) of air pollutant concentration levels can prevent deaths, regardless the initial level of exposure.

We observed evidence of a stronger effect of PM_{10} on natural-cause mortality in days characterized by higher temperature. This finding seems consistent with a previous investigation conducted on the entire territory of

Table 3 Attributable deaths in Milan for natural, respiratory and cardiovascular causes from 2003 to 2006: estimates and 90% confidence intervals calculated according to different counterfactual scenarios of PM_{10} reduction

Estimand	Natural Deaths		Respiratory Deaths		Cardiovascular Deaths	
	Estimate	90% CI	Estimate	90% CI	Estimate	90% CI
$AD(20)$	2537	(1273, 3655)	577	(226, 900)	1000	(464, 1488)
$AD(40)$	1157	(689, 1645)	312	(210, 418)	771	(580, 961)
$DAD(40)$	1857	(1479, 2233)	403	(293, 504)	1014	(815, 1221)
$DAD(50)$	1481	(1212, 1735)	325	(251, 398)	828	(690, 960)

$AD(20)$: deaths attributable to daily exposure levels above 20 $\mu g/m^3$, setting the counterfactual exactly to 20 $\mu g/m^3$.
$AD(40)$: deaths attributable to daily exposure levels above 40 $\mu g/m^3$, setting the counterfactual exactly to 40 $\mu g/m^3$.
$DAD(40)$: deaths attributable to daily exposure levels above 40 $\mu g/m^3$, sampling the counterfactuals from the exposure distribution below 40 $\mu g/m^3$.
$DAD(50)$: deaths attributable to daily exposure levels above 50 $\mu g/m^3$, sampling the counterfactuals from the exposure distribution below 50 $\mu g/m^3$

Table 4 Attributable deaths in Milan, Italy (2003-2006) calculated using different approaches

Source	Natural Deaths		Respiratory Deaths		Cardiovascular Deaths	
	Estimate	90% CI	Estimate	90% CI	Estimate	90% CI
GPS-*AD*(40)	1157	(689, 1645)	312	(210, 418)	771	(580, 961)
GPS-*DAD*(40)	1857	(1479, 2233)	403	(293, 504)	1014	(815, 1221)
PS matching 40	1079	(116, 2042)	305	(17, 593)	716	(117, 1315)
Regression 40	358	(156, 560)				

GPS-*AD*(40): deaths attributable to daily exposure levels above 40 μg/m^3, setting the counterfactual exactly to 40 μg/m^3.
GPS-*DAD*(40)): deaths attributable to daily exposure levels above 40 μg/m^3, sampling the counterfactuals from the exposure distribution below 40 μg/m^3.
PS matching 40: from Baccini et al. [25], deaths attributable to daily exposure levels above 40 μg/m^3, calculated according to a PS matching.
Regression 40: from Baccini et al. [37], deaths attributable to exceeding the limit of 40 μg/m^3 for the annual average level of exposure, calculated from a regression approach

the Lombardy region (for which Milan is the capital city), where higher effects of PM10 exposure on mortality were observed during summer [35].

In this paper, aimed at developing a new approach to estimate the dose-response curve and novel estimands for impact assessment, we focused on data on only one city. In the case of studies on multiple locations – which are common in this field – it is possible to use frequentist or Bayesian multivariate meta-analysis methods to combine the derivatives of the estimated aDRFs arising from several locations/cities [50]. We leave this extension to future work.

Comparison with previous results

The impacts estimated using the proposed approach partly confirmed the results reported in previous papers, obtained on the same data set with different methods. In Table 4 we report:

- the impact of the annual average PM$_{10}$ level exceeding 40 μg/m^3, calculated using the estimated coefficient from a Poisson regression, thus assuming the log-linearity of the effect [37] (*Regression 40* in Table 4);
- the impact of daily PM$_{10}$ levels above vs. below 40 μg/m^3, estimated using a PS matching, after dichotomizing the exposure at 40 μg/m^3 [25]. In this case, mortality in each "treated" day was compared to mortality in a matched day with similar characteristics (similar PS) but exposure under 40 μg/m^3 (*PS matching 40* in Table 4);
- our current findings *AD*(40) and *DAD*(40).

These estimates deserve some comments. First of all, the number of AD for natural causes estimated through the standard regression approach is far smaller than the other ones reported. This is due in part to the fact that, in the regression approach, the observed mortality was compared to the mortality that we would have observed during the study period, had the annual average level of exposure been set to 40 μg/m^{36}. This is different from

_____

[6]The number of attributable deaths due exceeding the threshold T for the annual average level of the air pollutant was calculated as

setting each counterfactual daily level of exposure above 40 μg/m^3 to exactly 40 μg/m^3 or to values below 40 μg/m^3, as in the present paper and in the PS matching approach [25]. In addition, we cannot exclude that the low estimated impacts can partly originate from the incorrect assumption of log-linearity on the exposure-response function.

Secondly, the estimated impacts using the GPS approach, especially *AD*(40), are consistent with the results obtained using PS matching. This is indicative of the substantial reproducibility of the results when applying procedures that, like these, do not force the log-linearity of the effect. At the same time, it also suggests that the simpler PS approach based on a binary version of the exposure can be considered a good "approximation" of the more complex method based on the GPS. However, the GPS approach, which estimates the attributable events according to a curve, returned more precise estimates than those obtained through PS matching in all the investigated outcomes, as shown by the narrower confidence intervals.

Finally, it is worth noticing that the impact *DAD*(40) estimated here is larger than the impact based on PS matching, even if both estimands seem to answer to the same research question: what would be the impact of a hypothetical intervention able to set daily exposure levels under 40 μg/m^3? The reason of this difference is that, as a matter of facts, the two estimands do not consider exactly the same counterfactual scenario. In fact, according to definition 4, *DAD*(40) and the PS matching estimand share the same threshold $z^{\star} = 40$, but set different $p^{\star}(z)$: for *DAD*(40) $p^{\star}(z)$ is the observed distribution of the daily exposures for all days with observed levels under the threshold, whereas for PS matching it is the distribution of the daily exposures among the matched controls [25]. Since in PS matching the control days with a level of exposure below 40 μg/m^3 are matched to the "treated"

_____

$y - y/exp(b(x - T))I(x > T)$, where y is the observed number of deaths during the study period, x is the annual average level of the air pollutant, b is the estimated regression coefficient expressing the air pollutant effect and $I(x > T)$ is an indicator function which is equal to 1 if $x > T$ and 0 elsewhere [37].

ones according to similarity in the observed characteristics, it is likely that the matched controls are those with a level of exposure closer to the threshold. For this reason, the impact estimated using PS matching was necessarily lower than $DAD(40)$ and close to $AD(40)$. This comparison highlights the need of providing a clear definition of the counterfactual scenario that is chosen to assess the impact.

Study limitations

Our study has some limitations. First of all, for low exposures, when the information is quite poor, a higher degree of smoothing for the bivariate spline of the outcome models would be needed, in order to reduce the sensitivity of the aDRFs to knots location. Adaptive splines could be used to reduce this problem [51, 52]. Second, the bootstrap procedure that we implemented in this analysis assumed independence between units. More complex approaches which account for the autocorrelation in the mortality time series could be used, such as block bootstrap or residual bootstrap methods [53]. Third, checking the balancing property of the GPS is crucial and not straightforward; we implemented a GPS blocking-based approach which employed t statistics to quantify unbalance [26]. As an alternative, standardized differences could be used as recently proposed [54].

It also seems appropriate to make a consideration regarding the positivity assumption. In this context, the positivity assumption would state that, for all possible strata defined by the covariates, it is possible to observe any level of exposure. As a matter of fact, even if in principle our approach does not require the validity of this assumption, its strong violation could challenge estimation, due to the presence of regions of the confounder space where inference would rely on extrapolation. A widely used method to address the violation of the positivity assumption consists in trimming the sample, which means excluding from the analysis the classes of units with limited variability in Z_i [55]. Although a slight evidence of violation of the positivity assumption arose in our data, we decided to perform the analysis on the entire data set anyway, without any kind of trimming. This choice was motivated by the need of quantifying the overall burden of mortality attributable to air pollution during the study period, without excluding any "treated" day.

Finally, formal statistical tests on the shape of the aDRF have not been developed within the GPS approach. As a consequence, the non-linearity of the aDRF could not be formally tested and our results on effect modification by temperature, although suggestive of a larger effect during the warm season, should be interpreted in a descriptive sense.

Conclusion

We adopted a method based on the GPS to obtain a semiparametric estimate of the average dose-response function describing the short-term effect of airborne particles on mortality and we defined novel estimands to assess the impact in terms of attributable deaths, according to the estimated curve. The estimands allow the specification of different counterfactual scenarios defined on the distribution of the daily air pollutant levels. This approach can be easily extended to other environmental epidemiology contexts. Its potential as a tool to investigate effect modification by temperature or other relevant factors deserves investigation.

Author details
[1]Department of Statistics, Computer Science, Applications, University of Florence, Viale Morgagni 59, 50134 Florence, Italy. [2]Department of Biostatistics, Yale School of Public Health, New Haven, CT, US. [3]Department of Clinical Sciences and Community Health, University of Milan and Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy.

References
1. Qian Z, He Q, Lin HM, et al. Association of daily cause-specific mortality with ambient particle air pollution in Wuhan China. Environ Res. 2007;105(3):380–9.
2. Samoli E, Analitis A, Touloumi G, et al. Estimating the exposure-response relationships between particulate matter and mortality within the APHEA multicity project. Environ Health Perspect. 2005;113(1):88–95. https://doi.org/10.1289/ehp.7387.

3.   Schwartz. J. Air pollution and daily mortality in Birmingham Alabama American. J Epidemiol. 1993;137(10):1136–47. https://doi.org/10.1093/oxfordjournals.aje.a116617.

4.   Smith RL, Spitzner D, Kim Y, Fuentes M. Threshold dependence of mortality effects for fine and coarse particles in Phoenix, Arizona. J Air Waste Manag Assoc. 2000;50(8):1367–79.

5.   Stafoggia M, Samoli E, Alessandrini E, et al. Short-term associations between fine and coarse particulate matter and hospitalizations in Southern Europe: results from the MED-PARTICLES project. Environ Health Perspect. 2013;121(9):1026–33. https://doi.org/10.1289/ehp.1206151.

6.   Daniels MJ, Dominici F, Zeger SL, Samet JM. The National Morbidity, Mortality, and Air Pollution Study. Part III: PM10 concentration-response curves and thresholds for the 20 largest US cities. Res Rep. 2004;(94 Pt 3): 1-21; discussion:3–30.

7.   Schwartz J, Zanobetti A. Using meta-smoothing to estimate dose-response trends across multiple studies, with application to air pollution and daily death. Epidemiology. 2000;11(6):666–72.

8.   Zanobetti A, Schwartz J. The effect of particulate air pollution on emergency admissions for myocardial infarction: a multicity case-crossover analysis. Environ Health Perspect. 2005;113(8):978–82. https://doi.org/10.1289/ehp.7550.

9.   Daniels MJ, Dominici F, Samet JM, Zeger SL. Estimating particulate matter-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest US cities. Am J Epidemiol. 2000;152(5):397–406. https://doi.org/10.1093/aje/152.5.397.

10.   Roberts S, Martin MA. The question of nonlinearity in the dose-response relation between particulate matter air pollution and mortality: can Akaike's Information Criterion be trusted to take the right turn? Am J Epidemiol. 2006;164(12):1242–50. https://doi.org/10.1093/aje/kwj335.

11.   Chen R, Yin P, Meng X, et al. Fine Particulate Air Pollution and Daily Mortality. A Nationwide Analysis in 272 Chinese Cities. Am J Respir Crit Care Med. 2017;196(1):73–81. https://doi.org/10.1164/rccm.201609-1862OC.

12.   Shen Y, Wu Y, Chen G, et al. Non-linear increase of respiratory diseases and their costs under severe air pollution. Environ Pollut. 2017;224:631–7. https://doi.org/10.1016/j.envpol.2017.02.047.

13.   Li T, Guo Y, Liu Y, et al. Estimating mortality burden attributable to short-term PM2.5 exposure: A national observational study in China. Environ Int. 2019;125:245–51. https://doi.org/10.1016/j.envint.2019.01.073.

14.   Zu K, Liu X, Shi L, et al. Concentration-response of short-term ozone exposure and hospital admissions for asthma in Texas. Environ Int. 2017;104:139–45. https://doi.org/10.1016/j.envint.2017.04.006.

15.   Liu C, Chen R, Sera F, et al. Ambient particulate air pollution and daily mortality in 652 cities. New England. J Med. 2019;381:705–15. https://doi.org/10.1056/NEJMoa1817364.

16.   United States Environmental Protection Agency. Regulatory Impact Analysis for the Final Revisions to the National Ambient Air Quality Standards for Particulate Matter. https://www3.epa.gov/ttn/ecas/docs/ria/naaqs-pm_ria_final_2012-12.pdf. Accessed August 2019.

17.   World Health Organization. Review of evidence on health aspects of air pollution - REVIHAAP Project. Copenhagen, Denmark: WHO; 2013.

18.   Gasparrini A, Leone M. Attributable risk from distributed lag models. BMC Med Res Method. 2014;14:55. https://doi.org/10.1186/1471-2288-14-55.

19.   Samoli E, Touloumi G, Zanobetti A, et al. Investigating the dose-response relation between air pollution and total mortality in the APHEA-2 multicity project. Occup Environ Med. 2003;60(12):977–82. https://doi.org/10.1136/oem.60.12.977.

20.   Vicedo-Cabrera AM, Sera F, Liu C, et al. Short term association between ozone and mortality: global two stage time series study in 406 locations in 20 countries. BMJ. 2020;10;368:m108. https://doi.org/10.1136/bmj.m108.

21.   Holland PW. Statistics and causal inference. J Am Stat Assoc. 1986;81(396): 945–60.

22.   Imbens GW, Rubin DB. Causal inference in statistics, social, and biomedical sciences. Cambridge: Cambridge University Press; 2015.

23.   Zigler CM, Dominici. F. Point: clarifying policy evidence with potential-outcomes thinking–beyond exposure-response estimation in air pollution epidemiology. Am J Epidemiol. 2014;180(12):1133–40. https://doi.org/10.1093/aje/kwu263.

24.   Schwartz J, Austin E, Bind MA, Zanobetti A, Koutrakis P. Estimating Causal Associations of Fine Particles With Daily Deaths in Boston. Am J Epidemiol. 2015;182(7):644–50. https://doi.org/10.1093/aje/kwv101.

25.   Baccini M, Mattei A, Mealli F, Bertazzi PA, Carugno M. Assessing the short term impact of air pollution on mortality: a matching approach. Environ Health Glob Access Sci Source. 2017;16(1):7. https://doi.org/10.1186/s12940-017-0215-7.

26.   Hirano K, Imbens GW. The propensity score with continuous treatments. In: Gelman A, Meng XL, editors. Applied Bayesian Modelling and Causal Inference from Missing Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family; 2004. p. 73–84.

27.   Imbens GW. The role of the propensity score in estimating dose-response functions. Biometrika. 2000;87(3):706–10.

28.   Bia M, Mattei A. A Stata package for the estimation of the dose-response function through adjustment for the generalized propensity score. Stata J. 2008;8(3):354–73.

29.   Flores CA, Flores-Lagunes A, Gonzalez A. Neumann TC. Estimating the effects of length of exposure to instruction in a training program: the case of job corps. Rev Econ Stat. 2012;94(1):153–71.

30.   Imai K, Van Dyk DA. Causal inference with general treatment regimes: Generalizing the propensity score. J Am Stat Assoc. 2004;99(467):854–66.

31.   Bia M, Flores CA, Flores-Lagunes A, Mattei A. A Stata package for the application of semiparametric estimators of dose-response functions. Stata J. 2014;14(3):580–604.

32.   Masson-Delmotte V, Zhai P, Pörtner H, et al. IPCC: Summary for Policymakers. In: Global Warming of 1.5 Celsius degrees. An IPCC Special Report on the impacts of global warming of 1.5 Celsius degrees above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty; 2018. https://bit.ly/2qusHsK. Accessed July 2019.

33.   Li J, Woodward A, Hou XY, et al. Modification of the effects of air pollutants on mortality by temperature: A systematic review and meta-analysis. Sci Total Environ. 2017;575:1556–70.

34.   Analitis A, De' Donato F, Scortichini M, et al. Synergistic Effects of Ambient Temperature and Air Pollution on Health in Europe: Results from the PHASE Project. Int J Environ Res Public Health. 2018;15(9):. https://doi.org/10.3390/ijerph15091856.

35.   Carugno M, Consonni D, Randi G, et al. Air pollution exposure, cause-specific deaths and hospitalizations in a highly polluted Italian region. Environ Res. 2016;147:415–24. https://doi.org/10.1016/j.envres.2016.03.003.

36.   Nawrot TS, Torfs R, Fierens F, et al. Stronger associations between daily mortality and fine particulate air pollution in summer than in winter: evidence from a heavily polluted region in western Europe. J Epidemiol Community Health. 2007;61(2):146–9. https://doi.org/10.1136/jech.2005.044263.

37.   Baccini M, Biggeri A, Grillo P, et al. Health impact assessment of fine particle pollution at the regional level. Am J Epidemiol. 2011;174(12): 1396–405. https://doi.org/10.1093/aje/kwr256.

38.   Baccini M, Grisotto L, Catelan D, et al. Commuting-adjusted short-term health impact assessment of airborne fine particles with uncertainty quantification via Monte Carlo simulation. Environ Health Perspect. 2015;123:27–33. https://doi.org/10.1289/ehp.1408218.

39.   INEMAR Emission Inventory, 2014. Emissions in the Milan Metropolitan Area - Public Review. ARPA Lombardia Settore Aria. https://bit.ly/2OgbKAM. Accessed July 2019.

40.   Katsouyanni K, Karakatsani A, Messari I, et al. Air pollution and cause specific mortality in Athens. J Epidemiol Community Health. 1990;44: 321–4. https://doi.org/10.1136/jech.44.4.321.

41.   Schwartz J, Dockery DW. Increased mortality in Philadelphia associated with daily air pollution concentrations. Am Rev Respir Dis. 1992;145(3): 600–4. https://doi.org/10.1164/ajrccm/145.3.600.

42.   Schwartz J, Spix C, Touloumi G, et al. Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. J Epidemiol Community Health. 1996;50(Suppl 1):S3–11.

43.   Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41–55.

44.   Wood SN. Generalized Additive Models: An Introduction with R (2nd edition). Boca Raton: Chapman and Hall, CRC Press; 2017.

45.   Efron B. Bootstrap methods: another look at the jackknife. Ann Stat. 1979;7:1–26.

46.   World Health Organization. WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide: global update 2005 - Summary of risk assessment. Geneva: WHO; 2006.

47. The European Parliament and of the Council Union. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. Offic J Eur Union. 2008;51(6):L152/1–L152/44. https://bit.ly/2ys7SSJ. Accessed July 2019.

48. Wu X, Braun D, Kioumourtzoglou M-A, Choirat C, Di Q, Dominici F. Causal inference in the context of an error prone exposure: air pollution and mortality. Ann Appl Stat. 2019;13(1):520–47.

49. United States Environmental Protection Agency. America's Children and the Environment, Third Edition - Criteria Air Pollutants. 2013. https://bit.ly/32NY7fi. Accessed July 2019.

50. Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. Stat Med. 2011;30(20):2481–98. https://doi.org/10.1002/sim.4172.

51. Krivobokova T, Crainiceanu CM, Kauermann G. Fast Adaptive Penalized Splines. J Comput Graph Stat. 2008;17(1):1–20.

52. Yang L, Hong Y. Adaptive penalized splines for data smoothing. Comput Stat Data Anal. 2017;180:70–83.

53. Kreiss J-P, Lahiri SN. Bootstrap methods for time series. In: Rao TS, Rao SS, Rao CR, editors. Time Series Analysis: Methods and Applications, Volume 30 (Handbook of statistics); 2012. p. 3–26.

54. Austin PC. Assessing covariate balance when using the generalized propensity score with quantitative or continuous exposures. Stat Methods Med Res. 2019;28(5):1365–77.

55. Petersen ML, Porter KE, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. Stat Methods Med Res. 2012;21(1):31–54.

56. Sandrini S, Fuzzi S, Piazzalunga A, et al. Spatial and seasonal variability of carbonaceous aerosol across Italy. Atmos Environ. 2014;99:587–98. https://doi.org/10.1016/j.atmosenv.2014.10.032.

57. Ruckerl R, Ibald-Mulli A, Koenig W, et al. Air pollution and markers of inflammation and coagulation in patients with coronary heart disease. Am J Respir Crit Care Med. 2006;173(4):432–41. https://doi.org/10.1164/rccm.200507-1123OC.

Publisher's Note