

# A CAUSAL EXPOSURE RESPONSE FUNCTION WITH LOCAL ADJUSTMENT FOR CONFOUNDING: ESTIMATING HEALTH EFFECTS OF EXPOSURE TO LOW LEVELS OF AMBIENT FINE PARTICULATE MATTER

BY GEORGIA PAPADOGEORGOU<sup>1</sup> AND FRANCESCA DOMINICI<sup>2</sup>

<sup>1</sup>*Department of Statistical Science, Duke University, [gp118@duke.edu](mailto:gp118@duke.edu)*

<sup>2</sup>*Department of Biostatistics, Harvard T.H. Chan School of Public Health, [fdominic@hsph.harvard.edu](mailto:fdominic@hsph.harvard.edu)*

In the last two decades ambient levels of air pollution have declined substantially. At the same time the Clean Air Act mandates that the National Ambient Air Quality Standards (NAAQS) must be routinely assessed to protect populations based on the latest science. Therefore, researchers should continue to address the following question: is exposure to levels of air pollution below the NAAQS harmful to human health? Furthermore, the contentious nature surrounding environmental regulations urges us to cast this question within a causal inference framework. Several parametric and semiparametric regression approaches have been used to estimate the exposure-response (ER) curve between long-term exposure to ambient air pollution concentrations and health outcomes. However, most of the existing approaches are not formulated within a formal framework for causal inference, adjust for the same set of potential confounders across all levels of exposure and do not account for model uncertainty regarding covariate selection and the shape of the ER.

In this paper we introduce a Bayesian framework for the estimation of a causal ER curve called LERCA (Local Exposure Response Confounding Adjustment), which: (a) allows for different confounders *and* different strength of confounding at the different exposure levels, and (b) propagates model uncertainty regarding confounders' selection and the shape of the ER. Importantly, LERCA provides a principled way of assessing the observed covariates' confounding importance at different exposure levels, providing researchers with important information regarding the set of variables to measure and adjust for in regression models. Using simulation studies, we show that state-of-the-art approaches perform poorly in estimating the ER curve in the presence of local confounding.

LERCA is used to evaluate the relationship between long-term exposure to ambient PM<sub>2.5</sub>, a key regulated pollutant, and cardiovascular hospitalizations for 5,362 zip codes in the continental U.S. and located near a pollution monitoring site, while adjusting for a potentially varying set of confounders across the exposure range. Our data set includes rich health, weather, demographic and pollution information for the years of 2011–2013. The estimated exposure-response curve is increasingly indicating that higher ambient concentrations lead to higher cardiovascular hospitalization rates, and ambient PM<sub>2.5</sub> was estimated to lead to an increase in cardiovascular hospitalization rates when focusing at the low-exposure range. Our results indicate that there is no threshold for the effect of PM<sub>2.5</sub> on cardiovascular hospitalizations.

**1. Introduction.** The Clean Air Act, one of the most comprehensive and expensive air quality regulations in the world, mandates that the National Ambient Air Quality Standards (NAAQS) are routinely reviewed. If evidence of the adverse health effects of exposure to

---

Received March 2019; revised January 2020.

*Key words and phrases.* Air pollution, cardiovascular hospitalizations, causal inference, exposure response function, local confounding, low-exposure levels, particulate matter.

ambient air pollution at levels below the NAAQS is established based on the peer reviewed literature, then the NAAQS must be lowered, even at the cost of hundreds of million of dollars. For that reason, researchers routinely address the following question: is exposure to levels of air pollution, even below the NAAQS, harmful to human health? With the next review of the NAAQS for fine particulate matter (PM<sub>2.5</sub>) scheduled to be completed by the end of the year 2020, the determination of whether exposure levels of PM<sub>2.5</sub> below the NAAQS is harmful to human health is subject to unprecedented level of scrutiny. More recently, because of the highly contentious nature surrounding air pollution regulations and the lowering of the NAAQS particularly, there is an increasing pressure to cast this question within a causal inference framework (Zigler and Dominici (2014)). The method in this paper is motivated by the need to address this critically important question by flexibly estimating an exposure response curve while reliably eliminating confounding bias, especially *at low levels* of exposure.

The literature on the harmful effects of air pollution is very extensive (see, e.g., Dominici et al. (2002), Eftim et al. (2008), Zanobetti and Schwartz (2007), Zeger et al. (2008), Crouse et al. (2015, 2016), Di et al. (2017a, 2017b), Berger et al. (2017), Lim et al. (2018), Makar et al. (2018)). However, significant methodological gaps remain in the context of estimating health effects at very low levels. Environmental research studying the health effects of exposure to low levels of ambient air pollution has either examined the relationship in the subset of the sample residing in areas with ambient concentrations below a prespecified threshold (Lee et al. (2016), Shi et al. (2016)), Di et al. (2017a, 2017b), Schwartz, Bind and Koutrakis (2017), Makar et al. (2018), Wang et al. (2018), Schwartz, Fong and Zanobetti (2018), or has employed regression approaches for ER estimation across the observed range of pollution concentrations (Daniels et al. (2000), Dominici et al. (2002), Schwartz, Laden and Zanobetti (2002), Bell, Peng and Dominici (2006), Hart et al. (2015), Thurston et al. (2016), Jerrett et al. (2017), Weichenthal et al. (2017), Lim et al. (2018)). In either case, confounding adjustment in air pollution studies is most often performed using either a prespecified set of covariates or a set of covariates which is decided upon using an ad hoc variable selection procedure. Such procedure is often based on the statistical significance of covariates' coefficients in an outcome regression or the change in the pollution concentration's coefficient in an outcome model including and excluding sets of covariates (Devries, Kriebel and Sama (2016), Pinault et al. (2016), Garcia et al. (2016), Weichenthal et al. (2017)).

Generally, regression and semiparametric modeling approaches for ER estimation, such as generalized linear models or generalized additive models (Hastie and Tibshirani (1986), Daniels et al. (2004), Shaddick et al. (2008), Shi et al. (2016), Dominici et al. (2002)), make the following assumptions: (1) the set of potential confounders that are included into the regression model among a potentially large set of available covariates is specified a priori; (2) uncertainty arising from the variable selection techniques is not accounted for; (3) the same potential confounders with constant confounding strength are considered when estimating the health effects across all exposure levels (we refer to this as *global confounding adjustment*); and (4) the shape of the ER function is modelled as a spline, a polynomial or linear with a threshold.

Even though ER estimation in air pollution research has mostly remained outside the potential outcome framework, there has been substantial work in ER estimation within the causal inference literature. Hirano and Imbens (2004) introduced the generalized propensity score (GPS) in order to adjust for confounding when estimating the effects of a continuous exposure. Flores, Flores-Lagunes and Gonzalez (2012) estimated a causal ER function employing a weighted locally linear regression with weights defined based on the GPS. Kennedy et al. (2017) introduced a doubly robust approach for estimating the causal ER function using flexible machine learning tools.

These approaches are very promising and manifest the growing interest in principled causal inference methods for continuous exposures. However, none of the existing approaches

explicitly accommodates that, in ER estimation and in contrast to binary treatments, confounding *might differ across levels of the exposure*. In fact, even though some of the approaches could be altered to allow for different set of confounders or different confounding strength across exposure levels, current implementations of causal methodology for ER estimation has assumed global confounding of preselected covariates. Furthermore, it is unclear how these approaches perform in the case of confounding that varies across exposure levels. To address this, confounding adjustment and confounder selection need to be meaningfully extended in the case of a continuous exposure to provide useful scientific guidance with regard to covariates' confounding importance *at different exposure levels*.

In our exploratory analyses (Section 2), we report that the relationship between ambient  $PM_{2.5}$  concentrations and the rate of hospitalization for cardiovascular diseases might be confounded by a *different set of covariates* at the low vs. at the high-exposure levels or by covariates with *different confounding strength*. We refer to this phenomenon as *local confounding*. We argue that, especially in the context of estimating causal effects at low levels, local confounding adjustment is deemed necessary.

To target local confounding, if exposure levels with different confounding were known, one could adopt a separate model at each level and adjust for *all* measured variables using one of the approaches described above. However, even if the number of covariates and local sample size rendered such approach computationally feasible, including unnecessary confounders in the regression model could lead to inefficient estimation of causal effects, especially at very low levels of exposure where data are sparse. Data driven methods to select a minimal necessary set of covariates to be included into an outcome model for estimation of causal effects of binary treatments have been proposed (de Luna, Waernbaum and Richardson (2011), Wang, Parmigiani and Dominici (2012), Wilson and Reich (2014)), but, to our knowledge, they have not been extended to the context of ER estimation with local confounding adjustment.

The goal of this paper is to overcome the challenges described above by introducing a Bayesian framework for the estimation of a causal ER curve called LERCA (Local Exposure Response Confounding Adjustment). We cast our approach within a causal inference framework by introducing the concept of *experiment configuration*  $\bar{s} = (s_0, s_1, \dots, s_{K+1})$ , where  $[s_{k-1}, s_k]$  denotes a specific range of exposure values. We use the term *experiment* to mimic the hypothetical assignment of a unit to exposure value within  $[s_{k-1}, s_k]$ . Within each experiment, that is, *locally* in the exposure range  $[s_{k-1}, s_k]$ , we assume that: (1) the ER is linear; (2) the potential confounders of the ER relationship are unknown but observed; and (3) the strength of the local confounding is also unknown. Across experiments we require that the ER is continuous at the points  $\bar{s}$ . Importantly, the internal points of the experiment configuration,  $s_1, s_2, \dots, s_K$ , are themselves unknown and have to be estimated from the data.

Our work contributes to various components in the literature. First, we contribute to the estimation of causal effects of continuous treatments by extending our understanding of confounding in these settings. Second, our work has connections to the literature on Bayesian free-knot splines (Denison, Mallick and Smith (1998), DiMatteo, Genovese and Kass (2001)). The location of the knots (internal points of the experiment configuration) is informed by both the ER fit and the necessity for local confounding adjustment. Lastly, our work contributes to the highly controversial and politically charged issue of estimating the causal effects of population exposure to low levels of ambient air pollution.

Even though our motivation and focus are the effects of air pollution, the statistical challenges related to ER estimation at low-exposure levels are common across many fields, such as toxicology (Scholze et al. (2001)) and clinical trials (Babb, Rogatko and Zacks (1998)). In fact, the methodology presented in this paper can be used to evaluate regulatory settings of potential harmful substances and can be routinely used to assess health effects of low

level exposures. Such applications include the effects of lead (Chiodo, Jacobson and Jacobson (2004), Jusko et al. (2008)), environmental contaminants (Van Der Oost, Beyer and Vermeulen (2003)), radiation (National Research Council (2006), Fazel et al. (2009)) and pesticides (Mackenzie Ross et al. (2010), Androutsopoulos et al. (2013)).

In Section 2 we introduce our motivating data set, discuss the difference between personal exposures and ambient concentrations in air pollution studies and illustrate that local confounding is likely to be present in our study. In Section 3 we introduce the notation and assumptions on which LERCA in Section 4 is based. In Section 5 we show through simulations that both off-the-shelf and state-of-the-art approaches for ER estimation perform poorly when local confounding is present, and we compare LERCA to alternatives in the presence of global confounding. Finally, in Section 6 we use LERCA to estimate the causal ER function relating ambient  $PM_{2.5}$  concentrations with log cardiovascular hospitalization rates in the Medicare population of 5,362 zip codes. Limitations and potential extensions are discussed in Section 7. Supplementary Material is available at Papadogeorgou and Dominici (2020).

**2. Data description, ambient concentrations, local confounding.** In this section we illustrate that, in our study, there might exist a different set of confounders at the low and the high levels of ambient pollution concentrations. LERCA is motivated to overcome this particular challenge.

*2.1. Data description.* We start by briefly describing our data set which is a collection of linked data from many sources. The unit of observation is the zip code  $i$ , with sample size  $N = 5,362$ . For each zip code, we acquired information on several potential confounders, denoted by  $C_{ij}$  for  $j = 1, 2, \dots, p$  and  $p = 27$ , capturing socioeconomic, demographic, climate and risk factor information. The full set of zip code level covariates are described in Table A.1. We calculate the outcome  $Y_i$  defined as log hospitalization rate for cardiovascular diseases (codes ICD-9 390 to 459) among Medicare beneficiaries residing in zip code  $i$  in the year 2013. Since Medicare beneficiaries are, in their plurality, individuals over the age of 65, our focus is on the health effect of  $PM_{2.5}$  on the elderly. For each zip code  $i$ , we assign exposure  $X_i$  defined as the average of daily levels of ambient  $PM_{2.5}$  concentrations for the years 2011 and 2012 recorded by EPA (U.S. Environmental Protection Agency) monitors within a six mile radius of zip code  $i$ 's centroid. The values of  $X_i$  range from 2.7 to  $18.3 \mu\text{g}/\text{m}^3$  (see Figure 1). We define  $X_i$  using the two years prior to the year whose outcome we analyze in order to respect the temporal ordering of treatment and outcome when drawing causal conclusions. Longer time lags could be considered, but in such settings our analysis would potentially be more susceptible to population mobility.

Since our definition of  $X_i$  requires the presence of an EPA monitor within six miles of a zip code's centroid, the zip codes included in our study are a subset of the full set of zip codes

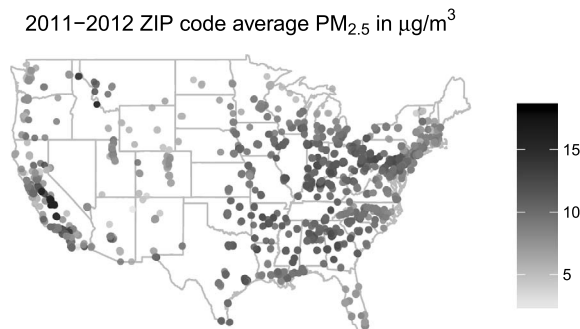


FIG. 1. Average levels of  $PM_{2.5}$  for the years 2011–2012 for each zip code  $i$  included into the analysis.

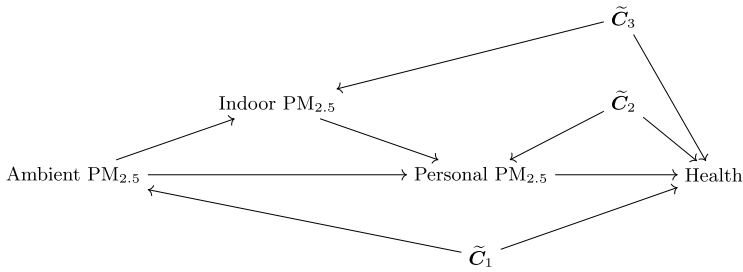


FIG. 2. DAG relating ambient PM<sub>2.5</sub> concentrations to personal exposures and health outcomes. Covariate sets  $\tilde{C}_1$ ,  $\tilde{C}_2$ ,  $\tilde{C}_3$  represent potentially different sets of confounders between the health outcome and ambient, personal or indoor PM<sub>2.5</sub> concentrations or exposures. Arrows represent potential (but not necessarily present) relationships.

in the continental U.S. Supplement A includes a detailed description of data linkage (EPA monitors, Medicare, others) and descriptive statistics across zip codes in the whole continental U.S. and only those included in our study. Excluded zip codes resemble, in general, those included in our analysis but are, perhaps, in more rural areas with lower population density and higher proportions of white population and unemployment.

**2.2. Ambient concentrations versus personal exposures.** In order to agree with existing causal inference literature for continuous treatments, we often refer to measurements  $X_i$  as zip code  $i$ 's *exposure* and a range of ambient pollution concentration as an *exposure level*. However, a zip code's measurement of ambient concentration  $X_i$  might be substantially different from the personal exposure of an individual residing in that zip code. Figure 2 shows a hypothetical DAG relating ambient and indoor pollution concentrations with individuals' personal exposures and health outcomes. Ambient pollution concentrations act on an individual's outcome only through the individuals' personal exposures.

In this paper we focus on estimating the causal effects of ambient PM<sub>2.5</sub> on cardiovascular health outcomes. That is, potentially, the most interesting question from a policy perspective, since policy regulations (and the NAAQS) are set based on the knowledge for the effect of ambient concentrations. The implications of using ambient concentrations instead of personal exposures to study the effect of pollution concentrations are discussed in Section 7.

**2.3. Potential presence of local confounding in our study.** In the case of binary treatments, whether a covariate acts as a confounder is often evaluated by checking whether there exists significant imbalance in the covariate distribution of the treated and control groups. For continuous exposures there is no direct counterpart to covariate balance since units are not separated into two groups. Instead, exploratory analyses for the presence of confounding are often based on covariates' strength in predicting the exposure through regression models (Imai and van Dyk (2004)). Then, a covariate's p-value in a model for the exposure is used to investigate whether it might be a confounding variable.

We use a related approach to illustrate the potential presence of local confounding in our data. We considered two subsets of zip codes: (1) zip codes with low ambient concentrations ( $<8 \mu\text{g}/\text{m}^3$ ; 817 observations), and (2) zip codes with high ambient concentrations ( $>11.5 \mu\text{g}/\text{m}^3$ ; 672 observations). Even though this definition of the low- and high-exposure levels is arbitrary for the purpose of our illustration, this choice ensures a similar number of observations and similar range of exposure values within the two levels. Within each exposure level *separately*, we considered a linear regression of ambient pollution concentration on each covariate and evaluated the covariate's predictive strength through its p-value.



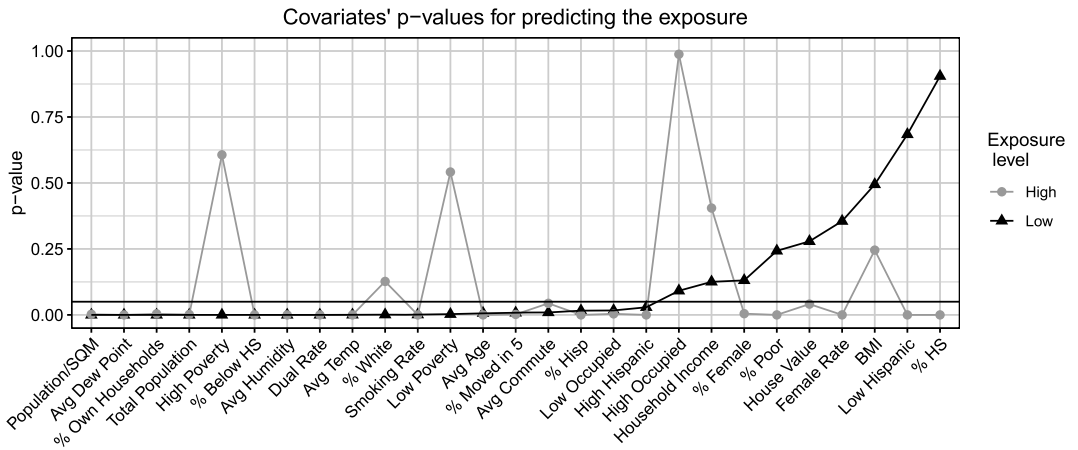


FIG. 3. Covariate p-values in predicting the exposure, separately at the low (dark grey:  $<8 \mu\text{g}/\text{m}^3$ ) and high (light grey:  $>11.5 \mu\text{g}/\text{m}^3$ ) exposure levels.

Figure 3 shows the p-value of the covariates in the  $p$  regression models and for the two exposure levels. We see that some variables, such as population density (Population/SQM) and the percentage of the population with less than a high school education (% Below HS), are predictive of ambient concentrations in both low- and high-exposure levels. However, other variables, such as the median household value (in logarithm—House Value), are only predictive of ambient pollution concentrations at the high-exposure levels. The opposite is true for the percentage of population that is white (% White). Such initial investigation indicates that different variables might act as predictors of the ambient exposure at different exposure levels.

In Supplement B, we show the estimated covariates’ coefficients whose p-values are shown in Figure 3. Since the two exposure levels are relatively balanced in terms of number of observations and range of exposure values, the magnitude of the p-values in Figure 3 is directly comparable to the magnitude of the estimated coefficients. This indicates that initial investigation of local confounding could be equivalently performed in terms of estimated coefficients or p-values.

In Supplement B we also consider a similar exploratory analysis to investigate which covariates are predictors of the health outcome at the low- and high-exposure levels separately. Combining the results presented there to the ones in Figure 3, there is evidence that the variables that confound the ER relationship might differ across levels of the exposure leading to *local confounding*. For example, the zip code median household value (House Value) is predictive of both ambient air pollution and cardiovascular hospitalization rates at the high-exposure levels but is not predictive of ambient air pollution at low-exposure levels. Additionally, there is indication that the percentage of the population with less than a high school degree (% Below HS) is a confounder at the low-exposure levels, whereas the same variable is not predictive of the health outcome at the high-exposure levels.

**3. Causal ER, the experiment configuration, and the local ignorability assumption.**

We follow the potential outcome framework (Hirano and Imbens (2004), Rubin (1974), Splawa-Neyman (1990)), and, under the stable unit value of treatment assumptions (SUTVA; no interference, no hidden versions of the treatment (Rubin (1980))), we use  $Y_i(x)$  to denote the potential outcome for observation  $i$  at exposure  $x \in \mathcal{X}$ , where  $\mathcal{X} \subset \mathbb{R}$  is the interval including all possible exposure values. Then,  $\{Y_i(x), x \in \mathcal{X}\}$  is unit  $i$ ’s ER curve, and  $\{\bar{Y}(x) = E[Y_i(x)], x \in \mathcal{X}\}$  is the population average ER curve.

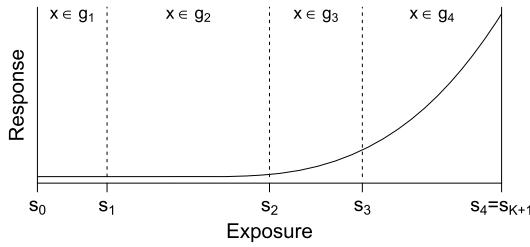


FIG. 4. ER curve with exposure range partitioned by  $\bar{s}$  in four experiments.

Assuming  $\bar{Y}(x)$  is differentiable as a function of  $x$ , we define the instantaneous causal effect as

$$\Delta(x) = \lim_{h \rightarrow 0} \frac{\bar{Y}(x + h) - \bar{Y}(x)}{h}.$$

A  $\Delta(x) \neq 0$  implies that variation in the exposure in a neighborhood of  $x$  has a causal effect on the expected outcome. We also define the population average causal effect of an exposure shift from  $x$  to  $x + \delta$ , as  $CE_\delta(x) = \bar{Y}(x + \delta) - \bar{Y}(x) = \int_x^{x+\delta} \Delta(t) dt$ . The observed outcome  $Y_i$  is equal to the potential outcome at the observed exposure  $Y_i(X_i)$ .

Under the weak ignorability assumption which states that the treatment is as if randomized conditional on observed covariates,  $X \perp\!\!\!\perp Y(x) | C$ , and every subject in the population can experience any  $x \in \mathcal{X}$ ,  $\bar{Y}(x)$  is identifiable using the observed data (Hirano and Imbens (2004)). Then, a minimal confounding adjustment set  $C^* \subseteq C$  is a set of covariates which satisfies  $X \perp\!\!\!\perp Y(x) | C^*$ , but  $X \not\perp\!\!\!\perp Y(x) | C^{**}$  for any  $C^{**}$  strict subset of  $C^*$  (de Luna, Waernbaum and Richardson (2011), Vansteelandt, Bekaert and Claeskens (2012), Wang, Parmigiani and Dominici (2012)).

In this paper we are interested in addressing the possibility that the minimal sufficient adjustment set  $C^*$  varies across exposure levels. We formalize this by introducing the *experiment configuration*. Let  $K$  denote a fixed positive integer and  $\min = \inf \mathcal{X}$  and  $\max = \sup \mathcal{X}$  denote the known and fixed minimum and maximum values of the exposure range  $\mathcal{X}$ . Then,  $\bar{s} = (s_0 = \min, s_1, s_2, \dots, s_K, s_{K+1} = \max)$  is the experiment configuration which defines a partition of the exposure range in  $K + 1$  experiments  $g_k = [s_{k-1}, s_k)$ . We use  $s$  to denote the internal points  $s_1, s_2, \dots, s_K$ . In Figure 4 a hypothetical exposure response function is plotted where  $\bar{s}$  defines a total of four experiments ( $K = 3$ ). Then,  $C_k^*$  is a minimal sufficient adjustment set in experiment  $k$  if it satisfies

$$(3.1) \quad X \perp\!\!\!\perp Y(x) | C_k^* \quad \text{for all } x \in g_k$$

and (3.1) does not hold for any strict subset of  $C_k^*$ . The sets  $C_k^*$  can be disjoint, identical or overlapping if the same variable is necessary for confounding adjustment in more than one experiment.

**4. ER estimation in the presence of local confounding.** Motivated by the evidence of local confounding between ambient  $PM_{2.5}$  concentrations and cardiovascular hospitalizations discussed in Section 2, we introduce LERCA: Local Exposure Response Confounding Adjustment. In order to build intuition, we do so for a fixed experiment configuration in Section 4.1. LERCA with unknown  $s$  is presented in Section 4.2. The choice of  $K$  is discussed in Section 4.4.

4.1. *Known experiment configuration.* Assume, for now, a known experiment configuration  $\bar{s}$ . Then, locally, that is, for  $x \in g_k = [s_{k-1}, s_k)$ , we assume the following pair of exposure and outcome models:

$$\begin{aligned}
 (4.1) \quad p(x|C = \mathbf{c}, x \in g_k) &= \phi\left(x; \delta_{k0}^X + \sum_{j=1}^p \alpha_{kj}^X \delta_{kj}^X c_j, \sigma_{k,X}^2\right), \\
 p(y|X = x, C = \mathbf{c}, x \in g_k) &= \phi\left(y; \delta_{k0}^Y + \beta_k(x - s_{k-1}) + \sum_{j=1}^p \alpha_{kj}^Y \delta_{kj}^Y c_j, \sigma_{k,Y}^2\right),
 \end{aligned}$$

where  $\phi(\cdot; \mu, \sigma^2)$  denotes the normal density with mean  $\mu$  and variance  $\sigma^2$  and  $\alpha_{kj}^X \in \{0, 1\}$  indicates that covariate  $C_j$  is included into the exposure model of the  $k$ th experiment, ( $\alpha_{kj}^X = 1$ ), or not, ( $\alpha_{kj}^X = 0$ ). The parameter  $\alpha_{kj}^Y$  has the same interpretation but for the outcome model. The parameter  $\beta_k$  denotes the instantaneous change in the expected outcome associated with a local variation in exposure for  $x \in g_k$  and adjusted for the  $C_j$ 's that have  $\alpha_{kj}^Y = 1$ . Even though all parameters depend on which covariates are included in the corresponding model, we do not explicitly state this dependence for notational simplicity. Model (4.1) allows for a different set of variables and variables' coefficients at the different experiments.

If a minimal confounding adjustment set for experiment  $k$  is included in the outcome model and the mean functional form is correctly specified,  $\beta_k$  is an unbiased estimator of the instantaneous effect  $\Delta(x)$ , for  $x \in g_k$ . Similarly, an unbiased estimator of  $\bar{Y}(x)$  is  $E_C\{E[Y_i|X = x, C]\}$  which can be estimated by taking the average over the units in our sample of the conditional expectation  $E[Y_i|X = x, C_i]$ .

In Section 4.1.3 we discuss how the prior distribution on the inclusion indicators is chosen to target confounding adjustment. In Section 4.1.4 we discuss prior specification for outcome model coefficients that ensures borrowing of information across experiments and ER continuity across the exposure range. But first we address two questions that naturally arise from the specification of model (4.1). First, we clarify the connection between LERCA and a model that specifies the ER relationship using linear splines in Section 4.1.1. Then, in Section 4.1.2 we discuss how LERCA compares to a model that is fit separately within each experiment  $g_k$ .

4.1.1. *Connection to linear splines.* In the outcome specification of model (4.1), the term  $\beta_k(x - s_{k-1})$  in the mean functional could be substituted by  $\beta_k x$ , and  $-\beta_k s_{k-1}$  could be absorbed in the intercept. However, specifying the model as to include  $\beta_k(x - s_{k-1})$  demonstrates the connection between model (4.1) and a model where the ER relationship is specified using linear splines with knots  $s$ . Furthermore, such specification significantly simplifies prior elicitation to ensure ER continuity (see Section 4.1.4) and posterior sampling satisfying the continuity condition (see Supplementary Material E).

Even though the outcome model in (4.1) resembles a linear splines model, there is a key distinction between the two models. In model (4.1) different experiments  $g_k$  are allowed to have a different slope for the exposure ( $\beta_k$ ), a different set of outcome predictors (covariates with  $\alpha_{kj}^Y = 1$ ) or the same set of predictors but with different coefficient ( $\delta_{kj}^Y$ ). Therefore, points  $s$  in (4.1) represent a change in the slope or a change in the outcome model covariate adjustment. On the other hand, a model that uses splines for the exposure-response relationship only allows  $\beta_k$  to vary with  $k$ . In this sense a splines model is a subcase of model (4.1) that for  $\alpha_{kj}^Y$  and  $\delta_{kj}^Y$  constant across  $k$ .

The assumption of local linearity (linear effect of the exposure on the outcome within each experiment) can lead to global nonlinearity and can be relaxed using higher-order splines.



However, for our study of the health effects of ambient air pollution at low concentrations previous research indicates that the relationship between air pollution and cardiovascular outcomes is linear (Thurston et al. (2016), Lim et al. (2018)) or supralinear (Crouse et al. (2015), Pinault et al. (2017)), situations to which our model can adjust.

4.1.2. *Connection to a separate model across experiments.* A natural question that arises from the LERCA model specification in (4.1) is how LERCA compares to fitting a separate outcome model within each experiment  $g_k$ . Doing so would still allow for different confounders and different confounding strength at different exposure levels.

However, a separate model within each experiment would not borrow any information across exposure levels and could estimate an ER that is not continuous at the points  $s$ . In contrast, LERCA borrows information across exposure levels by ensuring that the estimated ER is continuous everywhere (see Section 4.1.4). If higher order polynomials are used within each experiment, LERCA, similarly to splines, could be altered to accommodate higher-order smoothness across the exposure range.

4.1.3. *Prior distribution on inclusion indicators for confounding adjustment.* We build upon the work by Wang, Parmigiani and Dominici (2012), Wang et al. (2015) to assign an informative prior on covariates' local inclusion indicators  $(\alpha_{kj}^X, \alpha_{kj}^Y)$ . This prior choice ensures that model averaging assigns high-posterior weights to outcome models, including a minimal confounding adjustment set separately for each exposure range, and specifies

$$(4.2) \quad \frac{P(\alpha_{kj}^Y = 1 | \alpha_{kj}^X = 1)}{P(\alpha_{kj}^Y = 0 | \alpha_{kj}^X = 1)} = \omega \quad \text{where } \omega > 1, \text{ i.i.d. } \forall j, k.$$

By specifying (4.2), a variable  $C_j$  is assigned high prior probability to be included into the outcome model if it is also included in the exposure model ( $x \in g_k$  &  $\alpha_{kj}^X = 1$ ). Wang, Parmigiani and Dominici (2012) and Antonelli, Zigler and Dominici (2017) show that, for binary treatments, this informative prior leads to outcome models that include the minimal set of true confounders with higher posterior weights than model selection approaches that are based solely on the outcome model. In our context, this experiment-specific prior specification ensures that, locally, covariates in the minimal set  $C_k^*$  are included in the outcome model of experiment  $k$  with high posterior probability.

4.1.4. *Ensuring ER continuity.* In most applications it is expected that the causal ER relationship  $\bar{Y}(x)$  is continuous in  $x$ . Therefore, estimates of  $\bar{Y}(x)$ , in our case  $E_C\{E[Y|X = x, C]\}$ , should also be continuous. If the covariates  $C_j$  are centered and, under model (4.1), continuity of the estimated ER function is satisfied if

$$(4.3) \quad \lim_{x \rightarrow s_k^+} E_C\{E[Y|X = x, C]\} = \lim_{x \rightarrow s_k^-} E_C\{E[Y|X = x, C]\} \\ \iff \delta_{k0}^Y = \delta_{(k-1)0}^Y + \beta_{k-1}(s_k - s_{k-1}).$$

This is ensured by assuming a point-mass recursive prior on  $\delta_{k0}^Y, k \geq 2$ . Then, conditional on  $s$ , the outcome model intercept of experiment  $k \geq 2$  is a deterministic function of the outcome model intercept of the first experiment  $\delta_{10}^Y$ , and the slopes  $\beta_1, \beta_2, \dots, \beta_{k-1}$ . These parameters are assigned independent noninformative normal prior distributions.

4.1.5. *Prior distributions of the remaining coefficients.* Prior distributions on the remaining regression coefficients (exposure model coefficients, outcome model covariates' coefficients) and variance terms are chosen such that they lead to known forms of the full conditional posterior distributions to simplify sampling. We use independent noninformative inverse gamma prior distributions on  $\sigma_{k,X}^2, \sigma_{k,Y}^2$ . Noninformative normal prior is chosen for the exposure model intercepts  $\delta_{k0}^X$ . Conditional on the inclusion indicators, the prior on the regression coefficient  $\delta_{kj}^Y$  is a point mass at 0, or a noninformative normal distribution when  $\alpha_{kj}^Y$  is equal to 0 or 1, accordingly. Similarly, for the exposure model covariates' coefficients,  $\delta_{kj}^X$ . Default hyperparameter values are set to 0.001 for the inverse gamma distribution, (0, 100) for the mean and standard deviation of the normal distribution. Details on the prior specifications can be found in Supplement D.

4.2. *Unknown experiment configuration.* For a fixed experiment configuration  $\bar{s}$ , each experiment is treated separately in terms of confounder *selection and strength* of the confounding adjustment. However, the configuration itself is a key component of the fitted exposure response curve, and fixing it a priori could lead to bias and uncertainty underestimation. Instead, we assume that, a priori, the internal points of the experiment configuration  $s$  are distributed as the even-numbered order statistics of  $2K + 1$  samples from a uniform distribution on the interval  $(s_0, s_{K+1})$ . This prior choice of  $s$  discourages specifications of  $s$  that include values that are too close to each other (Green (1995)). The prior is augmented by indicators that consecutive points  $s_k, s_{k+1}$  cannot be closer than some distance  $d_k$ . Conditional on  $s$ , we follow the model specification and prior distributions described in Section 4.1.

4.3. *MCMC scheme and convergence diagnostics.* Markov chain Monte Carlo (MCMC) methods are used to acquire samples from the posterior distribution of model parameters. A detailed description of the MCMC scheme including computational challenges and contributions can be found in the Supplementary Material E. There, we also discuss MCMC convergence diagnostics based on the potential scale reduction factor (PSR; Gelman and Rubin (1992)) for quantities that do not directly depend on the experiment configuration.

4.4. *Number of points in the experiment configuration.* As presented previously, LERCA requires the specification of the internal number of points  $K$  in the experiment configuration. Since the number of parameters grows with  $K$ , possible values for  $K$  could be bounded by considering the maximum number of coefficients we are willing to entertain.

Cross-validation methods to choose values of tuning parameters are often infeasible in the Bayesian framework due to time and computational resources constraints. In a comprehensive review, Gelman, Hwang and Vehtari (2014) discusses methods of estimating the expected out of sample prediction error for Bayesian methods. The widely-applicable information criterion (WAIC; Watanabe (2010)) provides an estimate of the out-of-sample prediction error based on one MCMC run. It is defined as  $WAIC = -2(\text{lppd} - p_{WAIC})$ , where  $\text{lppd}$  and  $p_{WAIC}$  denote the log pointwise posterior predictive density and the penalty:

$$\begin{aligned} \text{lppd} &= \sum_{i=1}^n \log E_{\text{post}} p(x_i, y_i | \theta), \\ p_{WAIC} &= \sum_{i=1}^n \text{var}_{\text{post}}(\log p(x_i, y_i | \theta)). \end{aligned}$$

Here,  $\theta$  denotes the full vector of parameters, and  $E_{\text{post}}, \text{var}_{\text{post}}$  denote the posterior mean and variance.

In order to choose  $K$  for LERCA, LERCA is fit *once* for different values of  $K$ , and  $K$  is chosen as the value that minimizes the estimate of the WAIC.

**5. Simulation studies.** The main goal of our simulation study is to illustrate that local confounding is an important issue that both commonly-used and flexible approaches for ER estimation fail to adjust for, and they return biased results. The results from our simulation study indicate that methodology that directly accommodates local confounding is necessary in order to correctly estimate the causal effect of a continuous exposure.

Additionally, in Section 5.4 we discuss results from a simulation study under a generative model *without* local confounding. In this case traditional approaches and global confounding adjustment suffice for ER estimation, and the question is how comparably LERCA performs.

The approaches we considered are:

1. Generalized Additive Model (GAM): Regressing the outcome  $Y$  on flexible functions of the exposure  $X$  and all potential confounders (four degrees of freedom for each predictor).

2. Spline Model (SPLINE): Additive spline estimator described in [Bia et al. \(2014\)](#). The generalized propensity score (gps) is modelled as a linear regression on all covariates. The ER function is estimated using additive spline bases of the exposure and gps.

3. The Hirano and Imbens estimator ([Hirano and Imbens \(2004\)](#)) (HI-GPS): ER estimation is obtained by fitting an outcome regression model, including quadratic terms for both the exposure and the gps, and the exposure-gps interaction. The gps is estimated as in SPLINE.

4. Inverse Probability Weighting estimator (IPW): The generalized propensity score is used to weigh observations in an outcome regression model that includes linear and quadratic terms of exposure. The gps is estimated as in SPLINE.

5. The doubly-robust approach of [Kennedy et al. \(2017\)](#) (KENNEDY): The gps and outcome models are estimated using the Super Learner algorithm ([van der Laan, Polley and Hubbard \(2007\)](#)) combining the sample mean, linear regression with and without two-way interactions, generalized additive models, multivariate adaptive regression splines and random forests. Based on the gps and outcome model estimates, the pseudo outcome is calculated and is regressed on the exposure using kernel smoothing. This approach is chosen to represent state-of-the-art methods in ER estimation that are based on flexible, machine-learning and nonparametric approaches.

*5.1. Data generation with local confounding.* We generate data with exposure values which range from 0 to 10 and are uniformly distributed over the exposure range. Even though a uniform distribution is not accurate for the exposure variable in our study (ambient air pollution concentrations), we consider a uniformly distributed exposure to ensure that methods' performance is solely affected by the presence of local confounding and not by the presence of limited sample size at some exposure levels. We consider a quadratic ER and true experiment configuration  $\bar{s} = (0, 2, 4, 7, 10)$ . Table 1 summarizes which of the eight potential confounders are predictive of the exposure and/or the outcome within each experiment (correlations and regression coefficients are summarized in Table C.1). Note that in this data generating mechanism the minimal set of confounders vary across the four experiments. We simulate 400 data sets of 800 observations each. Details on the data generating mechanism are in Supplement F.

*5.2. Fitting the methods.* The different methods are fit using the `gam` and `causaldrf` R packages ([Hastie \(2017\)](#), [Schafer \(2015\)](#)), and the code available on [Kennedy et al. \(2017\)](#). LERCA is fit for  $K \in \{2, 3, 4\}$  and for each data set the results shown correspond to the  $K$  that minimized the WAIC.

Using each method, we estimate the population average ER curve  $\bar{Y}(x)$  over an equally spaced grid of points on the interval  $(0, 10)$ , and compare the root mean squared error (RMSE) as a function of  $x$ . We also assess whether LERCA can recover the correct location of the points  $s$ , identify the true confounders within each experiment and choose the correct value for  $K$ .

TABLE 1

Representation of which covariates are predictive of the exposure and/or the outcome within each experiment (denoted by a  $\checkmark$ ). Covariates with  $\checkmark$  in both models within the same experiment are local confounders.

Experiment	Model	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$	$C_8$
1	$X C$	$\checkmark$	$\checkmark$	$\checkmark$					
	$Y X, C$	$\checkmark$	$\checkmark$	$\checkmark$					
2	$X C$	$\checkmark$	$\checkmark$		$\checkmark$				
	$Y X, C$		$\checkmark$	$\checkmark$	$\checkmark$				
3	$X C$	$\checkmark$		$\checkmark$		$\checkmark$			
	$Y X, C$		$\checkmark$	$\checkmark$		$\checkmark$			
4	$X C$		$\checkmark$			$\checkmark$	$\checkmark$		
	$Y X, C$		$\checkmark$	$\checkmark$			$\checkmark$		

5.3. *Simulation results.* Figure 5 shows the estimated ER curves using the alternative methods. In Figure 6 we summarize the LERCA results including the estimated ER, the internal points of the experiment configuration and outcome model inclusion indicators of covariates  $C_1, C_4$  as a function of exposure  $x \in (0, 10)$ . We choose  $C_1$  and  $C_4$  because, in this data generating mechanism,  $C_1$  is a confounder in experiment 1 ( $x < 2$ ), and  $C_4$  is a confounder in experiment 2 only ( $2 < x < 4$ ). Grey lines correspond to results for individual data sets, whereas black solid lines correspond to averages across simulated data sets.

In Figure 5 we see that the alternative methods return biased results, especially at very low or very high levels of the exposure. These results indicate that neither commonly used nor flexible approaches utilizing machine learning tools appropriately accommodate local confounding adjustment for ER estimation. In terms of root MSE, LERCA was consistently lower than the alternative methods at low-exposure levels, all approaches performed similarly at middle-exposure levels and GAM slightly outperformed LERCA at high levels (Figure C.1). In the Supplementary Material C.2, we show that the relative performance of GAM and LERCA is reversed when the confounding structure is also reversed. These results indicate that local confounding is an issue across all exposure levels, and that, since the true confounding structure is never known for a real data set, LERCA should be preferred if local confounding is of concern.

As shown in Figure 6, even though the true ER is quadratic and LERCA is formulated as piecewise linear, LERCA is able to identify the correct shape of the exposure-response function. We find that using WAIC to choose the value of  $K$  led to choosing the correct

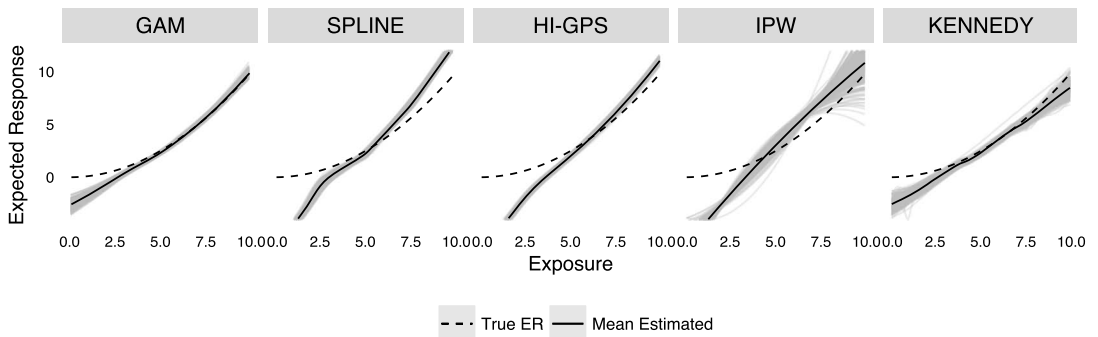


FIG. 5. The true mean ER function (dashed line), estimated ER functions from each simulated data set (gray), and the mean of the estimated ER functions (solid lines) using all alternative methods.

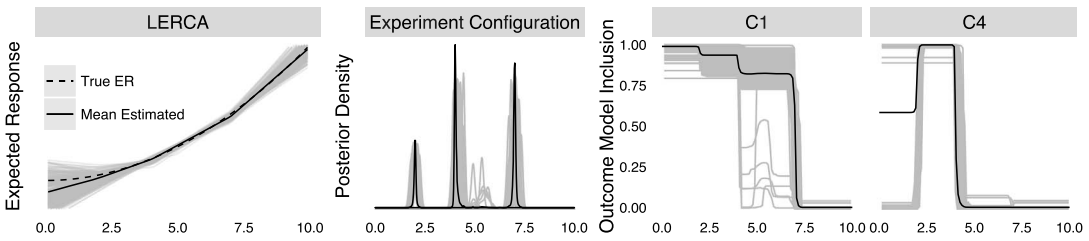


FIG. 6. LERCA results. (Left) Mean ER estimates. (Center) Posterior distribution of the internal locations  $s$ . (Right) Outcome model posterior inclusion probability of  $C_1$  and  $C_4$ . Gray lines correspond to simulated data sets separately, and black solid lines correspond to averages across data sets.

value of  $K = 3$  40% of the times and  $K = 2$  58% of the times, indicating that WAIC tends to overpenalize large values of  $K$ . Regardless, the correct internal locations  $s = \{2, 4, 7\}$  are located at the modes of the posterior distribution (second panel in Figure 6). By examining the posterior inclusion probabilities of  $C_1$ ,  $C_4$ , we observe that instrumental variables (e.g.,  $C_1$  in experiments 2 and 3) are often included in the outcome model. However, LERCA includes the minimal confounding set within each experiment with very high probability. On average (across the points in the exposure range and across all the simulated data sets), the minimal confounding set was included in the adjustment set 99% of the times (ranging from 89–100% across simulated data sets), indicating that the variables necessary for confounding adjustment are almost always included in the adjustment set. Lastly, the pointwise 95% and 50% credible intervals cover the true mean ER values 84% and 39% of the times, accordingly. The observed under-coverage is largely due to the underestimation of  $K$ .

5.4. *Simulation results in the absence of local confounding.* The previous generative scenario compared methods' performance in the presence of local confounding. In Supplement C.3, LERCA is compared to the alternative methods in the more traditional setting of global confounding, that is, in the setting more favorable to the other methods. In this context, LERCA with  $K = 3$  (fixed) performed similarly in terms of root MSE compared to GAM and Kennedy's doubly-robust estimator but better than the remaining alternative methods. These results indicate that LERCA offers a protection against bias arising from local confounding without sacrificing efficiency when local confounding is not present.

**6. Estimating the effect of ambient  $\text{PM}_{2.5}$  concentrations on zip code cardiovascular hospitalization rates.** We estimate the relationship between the average ambient  $\text{PM}_{2.5}$  concentrations for the years 2011–2012 and log cardiovascular hospitalization rates in 2013, using the data set introduced in Section 2 and allowing for local confounding adjustment. Here, a unit  $i$  from Section 3 corresponds to the areal unit of a zip code.

6.1. *Plausibility of the causal assumptions in our study.* The interpretation of estimated results as causal are bound by the plausibility of the causal assumptions within the study's setting. Here, we examine these assumptions in the evaluation of the causal relationship between ambient  $\text{PM}_{2.5}$  and cardiovascular hospitalization rates.

One assumption discussed in Section 3 is SUTVA which states that a zip code's potential outcomes are only a function of the zip code's own  $\text{PM}_{2.5}$  levels. If Medicare beneficiaries residing within a zip code travel outside of it, then other zip codes' ambient  $\text{PM}_{2.5}$  concentrations can affect the personal exposures of zip code  $i$ 's beneficiaries and, as a consequence, the zip code's hospitalization rates, invalidating SUTVA. This phenomenon is referred to in the literature as *interference*. Since  $\text{PM}_{2.5}$  concentrations in nearby zip codes are similar, and

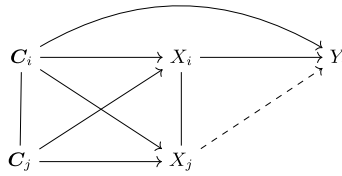


FIG. 7. Subgraph relating zip code  $i$  and  $j$ 's covariates and ambient  $PM_{2.5}$  concentrations to  $i$ 's potential outcomes. Arrow from  $X_j$  to  $Y_i$  is weaker than that from  $X_i$ .

Medicare beneficiaries are expected, at some level, to spend most of their time within a relatively close distance to their home, interference can be assumed to be limited (dashed arrow in Figure 7). When interference is limited, Sävje, Aronow and Hudgens (2018) showed that ignoring it returns estimates that are close to an average treatment effect.

The most commonly argued causal assumption is that of ignorability. In our setting, ignorability implies that, conditional on measured covariates, a zip code's "assignment" to a specific level of ambient  $PM_{2.5}$  does not depend on its potential outcomes (see (3.1)), and any zip code can experience any ambient  $PM_{2.5}$  within the observed range. One natural question is whether spatial correlation of ambient pollution concentrations invalidate ignorability, either through confounding or positivity. The no unmeasured confounding assumption is expected to hold if the set of measured covariates includes all confounders. In our study we might expect that the covariates of nearby zip codes (such as nearby weather conditions) affect a zip code's ambient  $PM_{2.5}$  concentrations (arrow from  $C_j$  to  $X_i$  in Figure 7). If, in addition, a zip code's outcome directly depends on the covariates of other zip codes (arrow from  $C_j$  to  $Y_i$ ), then  $C_j$  has to be included in the model for  $Y_i$ . We assume that such direct dependence does not exist in our study. For example, weather conditions near but not in zip code  $i$  only affect zip code  $i$ 's hospitalization rates through their effect on ambient  $PM_{2.5}$  concentrations. Even though ambient  $PM_{2.5}$  concentrations are spatially correlated, the positivity assumption requires that zip codes can experience any  $PM_{2.5}$  concentration level marginally, and the spatial correlation of  $PM_{2.5}$  does not further complicate the plausibility of the positivity assumption.

Lastly, the interpretation of our study results as causal is bound by the specification of the model in (4.1). If the mean functionals are not correctly specified, estimates of  $\beta_k$  can be biased for the causal effect of  $PM_{2.5}$  within that exposure level. Even though model (4.1) assumes independence of  $PM_{2.5}$  concentrations, the spatial dependence structure is not expected to affect estimation of the model's regression coefficients or variable selection.

The results presented next can only be interpreted as causal under the assumptions discussed here. If any of the assumptions is violated, the study results should be interpreted as associational.

6.2. Study results. We fit LERCA for  $K \in \{2, 3, \dots, 6\}$ , and we report the results for  $K = 3$  which corresponds to the model with the lowest WAIC. Figure 8 shows the posterior mean and the 95% credible intervals of the ER, the posterior distribution of the internal points of the experiment configuration and the posterior mean and 95% credible interval of  $\beta_k$  within each experiment. Positive values of  $\beta_k$  imply that an increase in ambient  $PM_{2.5}$  concentrations leads to an increase in hospitalization rates.

In Figure 8, we see that the estimated ER is supralinear with steeper incline at low concentrations. Examining the 95% credible intervals for  $\beta_k$ , there is evidence that an increase in  $PM_{2.5}$  at the low levels ( $\leq 9.9 \mu\text{g}/\text{m}^3$ ) leads to an increase in log hospitalization rates. However, 95% credible intervals for  $x \geq 9.9 \mu\text{g}/\text{m}^3$  include zero. Note that the current NAAQS for long-term exposure to ambient  $PM_{2.5}$  is equal to  $12 \mu\text{g}/\text{m}^3$ . These results indicate that there is no exposure threshold for the effect of  $PM_{2.5}$  on cardiovascular outcomes which



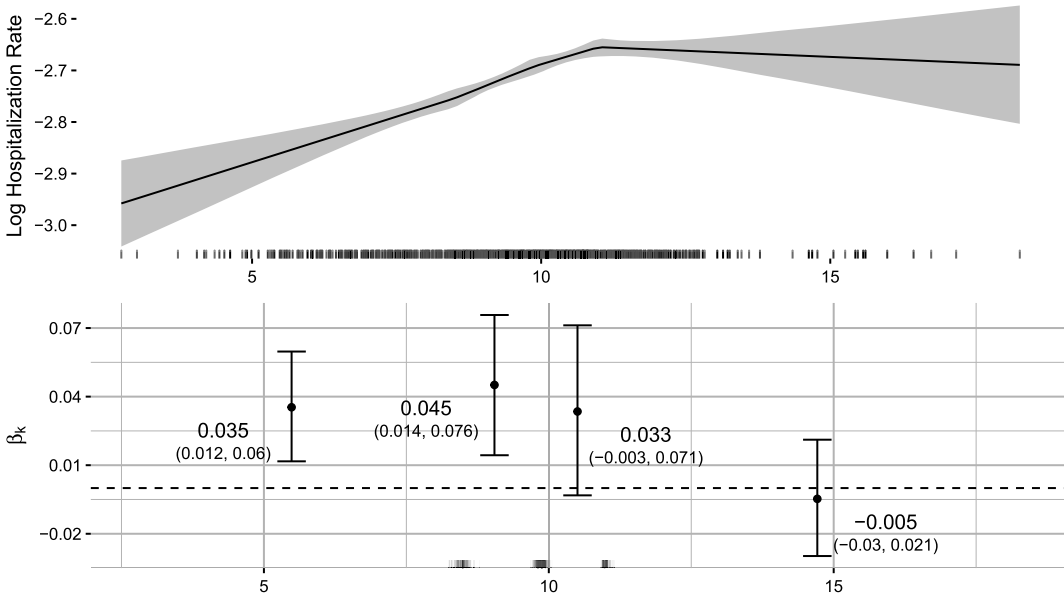


FIG. 8. Top: Mean ER curve of  $PM_{2.5}$  exposure (x-axis) and log all-cause cardiovascular hospitalization rates (y-axis)—solid line—with 95% pointwise credible intervals. The rug of points shows the distribution of observed  $PM_{2.5}$  values. Bottom: The posterior mean and 95% credible interval of the  $\beta$  coefficient within the four experiments. The rug of points shows the posterior distribution for  $s$  for  $K = 3$ .

means that reductions in ambient  $PM_{2.5}$  would lead to further health improvements, even at the low levels. These results are consistent with other epidemiological studies which have found that the strength of the association between  $PM_{2.5}$  and health outcomes is larger at low-concentration levels (Di et al. (2017b), Dominici et al. (2002), Shi et al. (2016)). Lastly, the posterior distribution of  $s$ , shows that observations below  $8 \mu g/m^3$  and over  $11.5 \mu g/m^3$  are always in the same experiment.

6.3. Variability of the covariates' posterior inclusion across  $PM_{2.5}$  concentration levels. We investigated whether local confounding was present by examining the variability of the covariates' inclusion probabilities in the exposure and outcome models as a function of  $PM_{2.5}$ . Figure 9 shows the posterior inclusion probabilities for three covariates as a function of  $PM_{2.5}$  providing a measure of the covariates' confounding importance across the  $PM_{2.5}$  concentration range.

The posterior inclusion probabilities vary substantially at different concentration levels, indicating that local confounding is likely to be present. In Figure 3 the exploratory analysis showed that the zip code median household value (House Value) was predictive of both

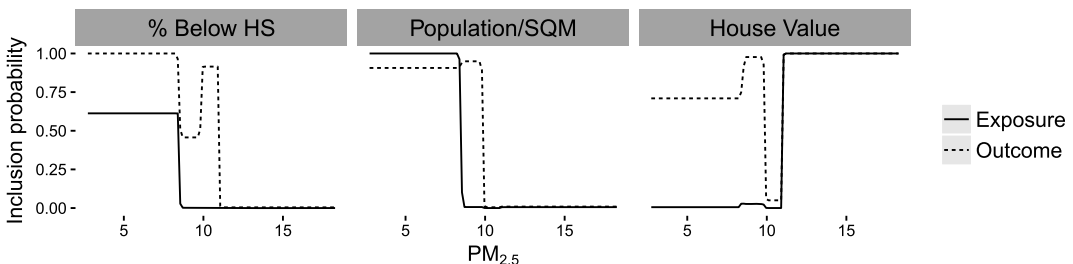


FIG. 9. Posterior inclusion probability of zip code population percentage with less than a high school education, population density, and median house value in the exposure and outcome model as a function of  $PM_{2.5}$ .

PM<sub>2.5</sub> and hospitalization rates at the high ambient concentration levels but only of the outcome at the low levels. The LERCA results in Figure 9 lead to the same conclusion. Similarly, the posterior inclusion probability for the variable representing the zip code's percentage of the population with less than a high school education (% Below HS) indicates that this variable is an important confounder only at the low levels, in accordance to the exploratory analysis. LERCA returns a similar conclusion about the variable representing population density (Population/SQM), in disagreement with the analysis in Section 2 which showed that population density was predictive of both PM<sub>2.5</sub> concentrations and the outcome at both low and high levels. Comparisons between the results in Figure 8 and the outcome model posterior inclusion probabilities were performed for all variables. LERCA tends to include in the outcome model a smaller number of variables than what one might have assumed based on the exploratory analysis. This is expected since LERCA considers the confounding importance of all variables simultaneously.

*6.4. Outcome model variability at low exposure levels.* With the focus of our study being the evaluation of the effect of ambient PM<sub>2.5</sub> at the low-concentration levels, we studied the interpretation of  $\beta_1$  across MCMC samples. Since the interpretation of  $\beta_1$  as a causal effect requires that a sufficient adjustment set is included in the outcome model, we examined the variability of the covariates' outcome model inclusion indicators within the low experiment across iterations of the MCMC.

Across MCMC samples, 174 combinations of the covariates were included in the outcome model (out of the  $2^{27}$  possible ones). Even though this is a large number of potential models, 51% of the posterior weight was given to the model with the following nine covariates: the zip code's median house value and percentage of the population with at most a high school education, as measured in the 2000 Census and its extrapolation between 2000 and 2013, the population rate that has been a smoker at some point in their lives, the zip code's population density, the average dew point, the average age of Medicare beneficiaries and the percentage of them that are women. We refer to this model as *Model 1*. The model with the second highest posterior probability included the same covariates, except for smoking rate, and accounted for 7% of the MCMC samples. Therefore, there is evidence that Model 1 outperformed the rest in confounding adjustment at low levels.

In order to evaluate the impact of model averaging on our final estimates, we compared the posterior distribution of  $\beta_1$  across all MCMC samples to its distribution based on the MCMC samples among which Model 1 was chosen. Across all samples,  $\beta_1$  was estimated to be equal to 0.035 with 95% credible interval 0.012–0.06 and posterior probability that it is greater than zero equal to 99.7%. Among posterior samples for which Model 1 was chosen,  $\beta_1$  was estimated to be 0.034 with 95% credible interval 0.011–0.056 and posterior probability that  $\beta_1$  is greater than zero also equal to 99.7%. The consistency of the Model 1 estimates and the model averaged estimates is an indication that model averaging, in this situation, did not lead to averaging over incompatible models.

**7. Discussion.** We have introduced an innovative Bayesian approach for flexible estimation of the ER curve in observational studies that has the following features: (1) it casts the formulation of the ER within a potential outcome framework and mimics several randomized experiments across exposure levels; (2) it uses the data to inform the experiment configuration; and, given the current experiment configuration, (3) allows for the possibility which is a reality in our study (Figure 3 and Figure 9) that different sets of covariates are confounders at different exposure levels; (4) allows for varying confounding effect across levels of the exposure; (5) performs local covariate selection to increase efficiency; (6) propagates model

uncertainty for the experiment configuration and covariate selection in the posterior inference on the whole ER curve, and; finally, (7) provides important scientific guidance related to which covariates are confounders at different exposure levels.

Although nonparametric and varying coefficient approaches (Hastie and Tibshirani (1993)) for ER estimation could, in theory, allow for differential confounding across different exposure levels, none of the existing methods for ER estimation explicitly accommodates local confounding nor provides guidance for which covariates are confounders of the effect of interest at different levels of the exposure. Furthermore, the use of nonparametric methods to estimate a generalized propensity score or to model the outcome of interest could prove unfruitful in situations where most of the available data are over a specific range of the exposure variable, the number of potential confounders is large and interest lies in the estimation of causal effects for a change in the exposure in the tails of the exposure distribution. In such situations LERCA provides a way to model the outcome acknowledging that the exposure-response relationship might be confounded by different covariates at different exposure levels. Lastly, it is worth noting that LERCA shall not be seen as a direct competitor to the approach by Kennedy et al. (2017). In fact, since the Super Learner algorithm combines different approaches for modeling the outcome, LERCA could be incorporated in the algorithm as an approach that allows for the presence of local confounding.

The main contribution of this paper is in addressing the issue of local confounding in ER estimation and in providing guidance of covariates' confounding importance at different exposure levels. In doing so, LERCA is based on several modeling decisions that can be easily altered. First, within each experiment and, thus, locally within a narrow exposure range, LERCA assumes linearity for both the outcome and exposure models. Local linearity could be relaxed by using higher order polynomials. Second, the informative prior on the inclusion indicators could lead to the inclusion of instrumental variables in the outcome model, which will not lead to bias but will decrease the efficiency of our estimators. In the study of air pollution, strong instrumental variables are not expected to be present. Alternative strategies for local confounder selection can be accommodated here, extending, for example, work by Cefalu et al. (2017), Wilson and Reich (2014) and Antonelli, Parmigiani and Dominici (2019). An interesting line of research is to explore LERCA's extensions to more flexible functional specifications and to evaluate the performance of different approaches to model averaging and model selection (via prior specifications or penalization techniques) for different confounding scenarios.

An alternative modification of LERCA could enforce that the ER curve is monotone, by assuming prior distributions on  $\beta_k$  that are left (or right)-truncated at zero. This modification could be of explicit interest for environmental and toxicological research and in studies of air pollution in particular where the ER relationship is often believed to be supralinear (Pinault et al. (2017), Vodonos, Awad and Schwartz (2018)). In risk assessment studies the shape of the ER can greatly affect conclusions (Pope et al. (2015)) and is often assumed to be linear, log-linear, log-log or power function with or without threshold (Devos et al. (2016), Burnett et al. (2014)). Nasari et al. (2016) proposed a class of models that can capture various ER shapes and are easy to implement in large data sets. Even though LERCA can capture effectively any ER shape, development of a faster and computationally efficient estimation procedure is required for very large data sets. Future work could focus on incorporating local confounding adjustment in air pollution analyses, including the whole United States and including zip codes that are not located near an air pollution monitor.

The results of our study are in agreement with a supralinear ER shape, indicating that there is a larger health impact of ambient air pollution concentrations at low exposure levels, and that, if the causal assumptions hold, lowering ambient  $PM_{2.5}$  concentrations would lead to a reduction in cardiovascular hospitalization rates. Even though our analysis addresses a key question in air pollution epidemiology, it is also met with its own challenges.

First, even though focusing on ambient pollution concentrations is important from a policy perspective (since regulations can more directly control ambient concentrations), the results of this study are not directly interpretable for evaluating the health effects of *personal* exposure to PM<sub>2.5</sub>. The relationship between ambient concentrations and personal exposures is complicated, with studies showing that there is substantial variability in personal exposures among individuals of similar ambient exposure concentrations (Dockery and Spengler (1981), Clayton et al. (1993)), largely due to the individuals' activities (Meng et al. (2009)). The correlation between ambient concentrations and personal exposures might differ by exposure levels if, for example, individuals residing in highly polluted areas are more likely to avoid outdoor activities. This further complicates interpreting the results of the relationship between ambient concentrations and health to personal exposures. Furthermore, variables that are expected to be confounders of personal exposures and health outcomes (such as an individual's smoking habits, variables  $C_2$  in Figure 2) are not necessarily confounders of ambient PM<sub>2.5</sub> concentrations and outcomes (variables  $C_1$  in Figure 2). Indication of confounding of the relationship between ambient concentrations and health outcomes by variables such as the median household value (see Figure 9) implies that these variables are not confounders in the classic sense (since they are not driving ambient PM<sub>2.5</sub> concentrations) but are correlated with variables that are. Therefore, a variable's confounding strength for ambient concentrations is not directly interpretable as its confounding strength for personal exposures.

Second, this analysis has used log event rates as the outcome of interest in a linear regression setting, with all zip codes contributing equally to the estimation of the models irrespective of their size. Linear regression for the analysis of rate data has been used in various settings, for example, in Chua et al. (2009), Joshua and Garber (1990), Mohamedshah, Paniati and Hobeika (1993), Wang, Parmigiani and Dominici (2012) and Liu-Smith et al. (2016). A Poisson regression model where the number of hospitalizations is the response variable and the Medicare population size within a zip code is the offset would be more in agreement with the literature on count outcomes (and within air pollution epidemiology specifically). However, extending local confounding adjustment to Poisson regression is computationally complicated: (a) enforcing ER continuity for Poisson regression is not straightforward since  $E_C\{E[Y|X = x, C]\}$  is not easily acquired (see Section 4.1.4), (b) posterior sampling of the experiment configuration and regression coefficient involves marginal densities of regression models, and (c) no conjugate prior distribution exists for regression coefficients in Poisson regression.

**Software.** An R package, which can be used to generate data with local confounding and fit LERCA, is available at <https://github.com/gpapadog/LERCA> and as Supplementary Material to this manuscript.

**Acknowledgments.** Funding for this work was provided by National Institutes of Health 1R01ES024332, 1R01GM111339, USEPA 83587201-0, and Health Effects Institute 4953-RFA14-3/16-4.

## SUPPLEMENTARY MATERIAL

**LERCA code** (DOI: [10.1214/20-AOAS1330SUPPA](https://doi.org/10.1214/20-AOAS1330SUPPA); .zip). We provide the R package used to generate data with local confounding and fit LERCA.

**Supplementary materials** (DOI: [10.1214/20-AOAS1330SUPPB](https://doi.org/10.1214/20-AOAS1330SUPPB); .pdf). This supplementary file includes (1) information on the creation of the data set, and included covariates, (2) further investigation of the presence of local confounding in our study, (3) simulation results under alternative data generative mechanisms, (4) information on the specification

of prior distributions for model parameters, (5) derivations for posterior sampling, including extensive information on sampling the experiment configuration and inclusion indicators to improve mixing, and (6) discussion on simulating data with local confounding.

## REFERENCES

- ANDROUTSOPOULOS, V. P., HERNANDEZ, A. F., LIESIVUORI, J. and TSATSAKIS, A. M. (2013). A mechanistic overview of health associated effects of low levels of organochlorine and organophosphorous pesticides. *Toxicology* **307** 89–94. <https://doi.org/10.1016/j.tox.2012.09.011>
- ANTONELLI, J., PARMIGIANI, G. and DOMINICI, F. (2019). High-dimensional confounding adjustment using continuous spike and slab priors. *Bayesian Anal.* **14** 825–848. MR3960772 <https://doi.org/10.1214/18-BA1131>
- ANTONELLI, J., ZIGLER, C. and DOMINICI, F. (2017). Guided Bayesian imputation to adjust for confounding when combining heterogeneous data sources in comparative effectiveness research. *Biostatistics* **18** 553–568. MR3799594 <https://doi.org/10.1093/biostatistics/kxx003>
- BABB, J., ROGATKO, A. and ZACKS, S. (1998). Cancer phase I clinical trials: Efficient dose escalation with overdose control. *Stat. Med.* **17** 1103–1120. [https://doi.org/10.1002/\(sici\)1097-0258\(19980530\)17:10<1103::aid-sim793>3.0.co;2-9](https://doi.org/10.1002/(sici)1097-0258(19980530)17:10<1103::aid-sim793>3.0.co;2-9)
- BELL, M. L., PENG, R. D. and DOMINICI, F. (2006). The exposure-response curve for ozone and risk of mortality and the adequacy of current ozone regulations. *Environ. Health Perspect.* **114** 532–536.
- BERGER, R. E., RAMASWAMI, R., SOLOMON, C. G. and DRAZEN, J. M. (2017). Air pollution still kills. *N. Engl. J. Med.* **376** 2591–2592.
- BIA, M., FLORES, C. A., FLORES-LAGUNES, A. and MATTEI, A. (2014). A Stata package for the application of semiparametric estimators of dose-response functions. *Stata J.* **14** 580–604.
- BURNETT, R. T., ARDEN POPE, C., EZZATI, M., OLIVES, C., LIM, S. S., MEHTA, S., SHIN, H. H., SINGH, G., HUBBELL, B. et al. (2014). An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environ. Health Perspect.* **122** 397–403. <https://doi.org/10.1289/ehp.1307049>
- CEFALU, M., DOMINICI, F., ARVOLD, N. and PARMIGIANI, G. (2017). Model averaged double robust estimation. *Biometrics* **73** 410–421. MR3665958 <https://doi.org/10.1111/biom.12622>
- CHIDO, L. M., JACOBSON, S. W. and JACOBSON, J. L. (2004). Neurodevelopmental effects of postnatal lead exposure at very low levels. *Neurotoxicol. Teratol.* **26** 359–371.
- CHUA, K. L., SOH, S. E., MA, S. and LEE, B. W. (2009). Pediatric asthma mortality and hospitalization trends across Asia Pacific. *World Allergy Organ. J.* **2** 77–82.
- CLAYTON, C. A., PERRITT, R. L., PELLIZZARI, E. D., THOMAS, K. W., WHITMORE, R. W., WALLACE, L. A., OZKAYNAK, H. and SPENGLER, J. D. (1993). Particle Total Exposure Assessment Methodology (PTEAM) study: Distributions of aerosol and elemental concentrations in personal, indoor, and outdoor air samples in a southern California community. *J. Expo. Anal. Environ. Epidemiol.* **3** 227–250.
- CROUSE, D. L., PETERS, P. A., HYSTAD, P., BROOK, J. R., VAN DONKELAAR, A., MARTIN, R. V., VILLENEUVE, P. J., JERRETT, M., GOLDBERG, M. S. et al. (2015). Ambient PM<sub>2.5</sub>, O<sub>3</sub>, and NO<sub>2</sub> exposures and associations with mortality over 16 years of follow-up in the Canadian Census Health and Environment Cohort (CanCHEC). *Environ. Health Perspect.* **123** 1180–1186. <https://doi.org/10.1289/ehp.1409276>
- CROUSE, D. L., PHILIP, S., DONKELAAR, A. V., MARTIN, R. V., JESSIMAN, B., PETERS, P. A., WEICHENTHAL, S., BROOK, J. R., HUBBELL, B. et al. (2016). A new method to jointly estimate the mortality risk of long-term exposure to fine particulate matter and its components. *Sci. Rep.* **6** Art. ID 18916. <https://doi.org/10.1038/srep18916>
- DANIELS, M. J., DOMINICI, F., SAMET, J. M. and ZEGER, S. L. (2000). 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.* **152** 397–406. <https://doi.org/10.1093/aje/152.5.397>
- DANIELS, M. J., DOMINICI, F., ZEGER, S. L. and SAMET, J. M. (2004). The national morbidity, mortality, and air pollution study. Part III: PM<sub>10</sub> concentration-response curves and thresholds for the 20 largest US cities. Research Report 94, Health Effects Institute.
- DE LUNA, X., WAERNBAUM, I. and RICHARDSON, T. S. (2011). Covariate selection for the nonparametric estimation of an average treatment effect. *Biometrika* **98** 861–875. MR2860329 <https://doi.org/10.1093/biomet/asr041>
- DENISON, D. G. T., MALLICK, B. K. and SMITH, A. F. M. (1998). Automatic Bayesian curve fitting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 333–350. MR1616029 <https://doi.org/10.1111/1467-9868.00128>
- DEVOS, S., COX, B., LIER, T. V., NAWROT, T. S. and PUTMAN, K. (2016). Effect of the shape of the exposure-response function on estimated hospital costs in a study on non-elective pneumonia hospitalizations related to particulate matter. *Environ. Int.* **94** 525–530.



- DEVRIES, R., KRIEBEL, D. and SAMA, S. (2016). Low level air pollution and exacerbation of existing COPD: A case crossover analysis. *Environ. Health* **15** Art. ID 98. <https://doi.org/10.1186/s12940-016-0179-z>
- DI, Q., DAI, L., WANG, Y., ZANOBETTI, A., CHOIRAT, C., SCHWARTZ, J. D. and DOMINICI, F. (2017a). Association of short-term exposure to air pollution with mortality in older adults. *J. Am. Med. Assoc.* **318** 2446–2456. <https://doi.org/10.1001/jama.2017.17923>. Association
- DI, Q., WANG, Y., ZABONETTI, A., WANG, Y., KOUTRAKIS, P., CHOIRAT, C., DOMINICI, F. and SCHWARTZ, J. D. (2017b). Air pollution and mortality in the medicare population. *N. Engl. J. Med.* **376** 2513–2522. <https://doi.org/10.1056/NEJMoa1702747>. Air
- DIMATTEO, I., GENOVESE, C. R. and KASS, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika* **88** 1055–1071. MR1872219 <https://doi.org/10.1093/biomet/88.4.1055>
- DOCKERY, D. W. and SPENGLER, J. D. (1981). Personal exposure to respirable particulates and sulfates. *J. Air Pollut. Control Assoc.* **31** 153–159. <https://doi.org/10.1080/00022470.1981.10465205>
- DOMINICI, F., DANIELS, M., ZEGER, S. L. and SAMET, J. M. (2002). Air pollution and mortality: Estimating regional and national dose-response relationships. *J. Amer. Statist. Assoc.* **97** 100–111. MR1963390 <https://doi.org/10.1198/016214502753479266>
- EFTIM, S. E., SAMET, J. M., JANES, H., MCDERMOTT, A. and DOMINICI, F. (2008). Fine particulate matter and mortality: A comparison of the six cities and American Cancer Society cohorts with a medicare cohort. *Epidemiology* **19** 209–216.
- FAZEL, R., KRUMHOLZ, H. M., WANG, Y., ROSS, J. S., CHEN, J., TING, H. H., SHAH, N. D., NASIR, K., EINSTEIN, A. J. et al. (2009). Exposure to low-dose ionizing radiation from medical imaging procedures. *N. Engl. J. Med.* **361** 849–857.
- GARCIA, C. A., YAP, P., PARK, H. and WELLER, B. L. (2016). Association of long-term PM<sub>2.5</sub> exposure with mortality using different air pollution exposure models: Impacts in rural and urban California. *Int. J. Environ. Health Res.* **26** 145–157. <https://doi.org/10.1080/09603123.2015.1061113>
- GELMAN, A., HWANG, J. and VEHTARI, A. (2014). Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **24** 997–1016. MR3253850 <https://doi.org/10.1007/s11222-013-9416-2>
- GELMAN, A. and RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–511.
- FLORES, C. A., NEUMANN, A., GONZALEZ, T. C. and GONZALEZ, A. (2012). Estimating the effects of length of exposure to instruction in a training program: The case of job corps. *Rev. Econ. Stat.* **94** 153–171.
- GREEN, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–732. MR1380810 <https://doi.org/10.1093/biomet/82.4.711>
- HART, J. E., LIAO, X., HONG, B., PUETT, R. C., YANOSKY, J. D., SUH, H., KIOUMOURTZOGLU, M., SPIEGELMAN, D. and LADEN, F. (2015). The association of long-term exposure to PM<sub>2.5</sub> on all-cause mortality in the Nurses' Health Study and the impact of measurement-error correction. *Environ. Health* **14** Art. ID 38. <https://doi.org/10.1186/s12940-015-0027-6>
- HASTIE, T. (2017). gam: Generalized additive models.
- HASTIE, T. and TIBSHIRANI, R. (1986). Generalized additive models. *Statist. Sci.* **1** 297–318. MR0858512
- HASTIE, T. and TIBSHIRANI, R. (1993). Varying-coefficient models. *J. Roy. Statist. Soc. Ser. B* **55** 757–796. MR1229881
- HIRANO, K. and IMBENS, G. W. (2004). The propensity score with continuous treatments. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. Wiley Ser. Probab. Stat. 73–84. Wiley, Chichester. MR2134803 <https://doi.org/10.1002/0470090456.ch7>
- IMAI, K. and VAN DYK, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *J. Amer. Statist. Assoc.* **99** 854–866. MR2090918 <https://doi.org/10.1198/016214504000001187>
- JERRETT, M., TURNER, M. C., BECKERMAN, B. S., III, C. A. P. and DONKELAAR, A. V. (2017). Comparing the health effects of ambient particulate matter estimated using ground-based versus remote sensing exposure estimates. *Environ. Health Perspect.* **125** 552–559.
- JOSHUA, S. C. and GARBER, N. J. (1990). Estimating truck accident rate and involvements using linear and Poisson regression models. *Transp. Plann. Technol.* **15** 41–58.
- JUSKO, T. A., HENDERSON, C. R., LANPHEAR, B. P., CORY-SLECHTA, D. A., PARSONS, P. J. and CANFIELD, R. L. (2008). Blood lead concentrations < 10  $\mu\text{g}/\text{dL}$  and child intelligence at 6 years of age. *Environ. Health Perspect.* **116** 243–248.
- KENNEDY, E. H., MA, Z., MCHUGH, M. D. and SMALL, D. S. (2017). Non-parametric methods for doubly robust estimation of continuous treatment effects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1229–1245. MR3689316 <https://doi.org/10.1111/rssb.12212>
- LEE, M., KOUTRAKIS, P., COULL, B. A., KLOOG, I. and SCHWARTZ, J. D. (2016). Acute effect of fine particulate matter on mortality in three southeastern states 2007–2011. *J. Expo. Sci. Environ. Epidemiol.* **26** 173–179. <https://doi.org/10.1038/jes.2015.47>. Acute



- LIM, C. C., HAYES, R. B., AHN, J., SHAO, Y., SILVERMAN, D. T., JONES, R. R., GARCIA, C. and THURSTON, G. D. (2018). Association between long-term exposure to ambient air pollution and diabetes mortality in the US. *Environ. Res.* **165** 330–336. <https://doi.org/10.1016/j.envres.2018.04.011>
- LIU-SMITH, F., FARHAT, A. M., ARCE, A., ZIOGAS, A., TAYLOR, T., WANG, Z., YOURK, V., LIU, J., WU, J. et al. (2016). Gender differences in the association between the incidence of cutaneous melanoma and geographic UV exposure. *J. Am. Acad. Dermatol.* **76** 499–505. <https://doi.org/10.1016/j.jaad.2016.08.027>
- MAKAR, M., ANTONELLI, J., DI, Q., CUTLER, D. and SCHWARTZ, J. (2018). Estimating the causal effect of fine particulate matter levels on death and hospitalization: Are levels below the safety standards harmful?. *Epidemiology* **28** 627–634. <https://doi.org/10.1097/EDE.0000000000000690>. Estimating
- MENG, Q. Y., SPECTOR, D., COLOME, S. and TURPIN, B. (2009). Determinants of indoor and personal exposure to PM(2.5) of indoor and outdoor origin during the RIOPA study. *Atmos. Environ.* **43** 5750–5758. <https://doi.org/10.1016/j.atmosenv.2009.07.066>
- MOHAMEDSHAH, Y. M., PANIATI, J. F. and HOBEIKA, A. G. (1993). Truck accident models for interstates and two-lane rural roads. Transportation Research Record 1407, TRB, National Research Council.
- NASARI, M. M., SZYSZKOWICZ, M., CHEN, H., CROUSE, D., TURNER, M. C., JERRETT, M., POPE, C. A., HUBBELL, B., FANN, N. et al. (2016). A class of non-linear exposure-response models suitable for health impact assessment applicable to large cohort studies of ambient air pollution. *Air Qual. Atmos. Health* **9** 961–972. <https://doi.org/10.1007/s11869-016-0398-z>
- NATIONAL RESEARCH COUNCIL (2006). *Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII Phase 2*. The National Academies Press, Washington, DC.
- PAPADOGEORGOU, G. and DOMINICI, F. (2020). Supplement to “A causal exposure response function with local adjustment for confounding: Estimating health effects of exposure to low levels of ambient fine particulate matter.” <https://doi.org/10.1214/20-AOAS1330SUPPA>, <https://doi.org/10.1214/20-AOAS1330SUPPB>
- PINAULT, L., TJEPKEMA, M., CROUSE, D. L., WEICHENTHAL, S., VAN DONKELAAR, A., MARTIN, R. V., BRAUER, M., CHEN, H. and BURNETT, R. T. (2016). Risk estimates of mortality attributed to low concentrations of ambient fine particulate matter in the Canadian community health survey cohort. *Environ. Health* **15** Art. ID 18. <https://doi.org/10.1186/s12940-016-0111-6>
- PINAULT, L. L., WEICHENTHAL, S., CROUSE, D. L., BRAUER, M., ERICKSON, A., VAN DONKELAAR, A., MARTIN, R. V., HYSTAD, P., CHEN, H. et al. (2017). Associations between fine particulate matter and mortality in the 2001 Canadian Census Health and Environment Cohort. *Environ. Res.* **159** 406–415. <https://doi.org/10.1016/J.ENVRES.2017.08.037>
- POPE, C. A., CROPPER, M., COGGINS, J. and COHEN, A. (2015). Health benefits of air pollution abatement policy: Role of the shape of the concentration-response function. *J. Air Waste Manage. Assoc.* **65** 516–522. <https://doi.org/10.1080/10962247.2014.993004>
- MACKENZIE ROSS, S. J., BREWIN, C. R., CURRAN, H. V., FURLONG, C. E., ABRAHAM-SMITH, K. M. and HARRISON, V. (2010). Neuropsychological and psychiatric functioning in sheep farmers exposed to low levels of organophosphate pesticides. *Neurotoxicol. Teratol.* **32** 452–459. <https://doi.org/10.1016/j.ntt.2010.03.004>
- RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688–701.
- RUBIN, D. B. (1980). Randomization analysis of experimental data: The Fisher randomization test comment. *J. Amer. Statist. Assoc.* **75** 591–593.
- SÄVJE, F., ARONOW, P. M. and HUDGENS, M. G. (2018). Average treatment effects in the presence of unknown interference.
- SCHAFFER, J. (2015). causalDRF: Tools for estimating causal dose response functions.
- SCHOLZE, M., BOEDEKER, W., FAUST, M., BACKHAUS, T., ALTENBURGER, R. and HORST, L. (2001). A general best-fit method for concentration-response curves and the estimation of low-effect concentrations. *Environ. Toxicol. Chem.* **20** 448–457.
- SCHWARTZ, J., BIND, M.-A. and KOUTRAKIS, P. (2017). Estimating causal effects of local air pollution on daily deaths: Effect of low levels. *Environ. Health Perspect.* **125** 23–29. <https://doi.org/10.1289/EHP232>
- SCHWARTZ, J., FONG, K. and ZANOBETTI, A. (2018). A national multicity analysis of the causal effect of local pollution, NO<sub>2</sub>, and PM<sub>2.5</sub> on mortality. *Environ. Health Perspect.* **126** Art. ID 087004.
- SCHWARTZ, J., LADEN, F. and ZANOBETTI, A. (2002). The concentration-response relation between PM<sub>2.5</sub> and daily deaths. *Environ. Health Perspect.* **110** 1025–1029.
- SHADDICK, G., LEE, D., ZIDEK, J. V. and SALWAY, R. (2008). Estimating exposure response functions using ambient pollution concentrations. *Ann. Appl. Stat.* **2** 1249–1270. MR2655658 <https://doi.org/10.1214/08-AOAS177>
- SHI, L., ZANOBETTI, A., KLOOG, I., COULL, B. A., KOUTRAKIS, P., MELLY, S. J. and SCHWARTZ, J. D. (2016). Low-concentration PM<sub>2.5</sub> and mortality: Estimating acute and chronic effects in a population-based study. *Environ. Health Perspect.* **124** 46–52.

- NEYMAN, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. Sci.* **5** 465–480.
- THURSTON, G. D., AHN, J., CROMAR, K. R., SHAO, Y., REYNOLDS, H. R., JERRETT, M., LIM, C. C., SHANLEY, R., PARK, Y. et al. (2016). Ambient particulate matter air pollution exposure and mortality in the NIH–AARP diet and health cohort. *Environ. Health Perspect.* **124** 484–490. <https://doi.org/10.1289/ehp.1509676>
- VAN DER LAAN, M. J., POLLEY, E. C. and HUBBARD, A. E. (2007). Super learner. *Stat. Appl. Genet. Mol. Biol.* **6** Art. ID 25. MR2349918 <https://doi.org/10.2202/1544-6115.1309>
- VAN DER OOST, R., BEYER, J. and VERMEULEN, N. P. E. (2003). Fish bioaccumulation and biomarkers in environmental risk assessment: A review. *Environ. Toxicol. Pharmacol.* **13** 57–149.
- VANSTEELANDT, S., BEKAERT, M. and CLAESKENS, G. (2012). On model selection and model misspecification in causal inference. *Stat. Methods Med. Res.* **21** 7–30. MR2867536 <https://doi.org/10.1177/0962280210387717>
- VODONOS, A., AWAD, Y. A. and SCHWARTZ, J. (2018). The concentration-response between long-term PM<sub>2.5</sub> exposure and mortality; a meta-regression approach. *Environ. Res.* **166** 677–689. <https://doi.org/10.1016/j.envres.2018.06.021>
- WANG, C., PARMIGIANI, G. and DOMINICI, F. (2012). Bayesian effect estimation accounting for adjustment uncertainty. *Biometrics* **68** 661–686. MR3055168 <https://doi.org/10.1111/j.1541-0420.2011.01731.x>
- WANG, C., DOMINICI, F., PARMIGIANI, G. and ZIGLER, C. M. (2015). Accounting for uncertainty in confounder and effect modifier selection when estimating average causal effects in generalized linear models. *Biometrics* **71** 654–665. MR3402601 <https://doi.org/10.1111/biom.12315>
- WANG, Y., SHI, L., LEE, M., LIU, P., DI, Q., ZABONETTI, A. and SCHWARTZ, J. D. (2018). Long-term exposure to PM<sub>2.5</sub> and mortality among older adults in the southeastern US. *Epidemiology* **28** 207–214. <https://doi.org/10.1097/EDE.0000000000000614>. Long-term
- WATANABE, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11** 3571–3594. MR2756194
- WEICHTHAL, S., KULKA, R., LAVIGNE, E., RIJSWIJK, D. V., BRAUER, M., VILLENEUVE, P. J., STIEB, D., JOSEPH, L. and BURNETT, R. T. (2017). Biomass burning as a source of ambient fine particulate air pollution and acute myocardial infarction. *Epidemiology* **28** 329–337. <https://doi.org/10.1097/EDE.0000000000000636>
- WILSON, A. and REICH, B. J. (2014). Confounder selection via penalized credible regions. *Biometrics* **70** 852–861. MR3295746 <https://doi.org/10.1111/biom.12203>
- ZANOBETTI, A. and SCHWARTZ, J. (2007). Particulate air pollution, progression, and survival after myocardial infarction. *Environ. Health Perspect.* **115** 769–775. <https://doi.org/10.1289/ehp.9201>
- ZEGER, S. L., DOMINICI, F., MCDERMOTT, A. and SAMET, J. M. (2008). Mortality in the Medicare population and chronic exposure to fine particulate air pollution in urban centers (2000–2005). *Environ. Health Perspect.* **116** 1614–1619. <https://doi.org/10.1289/ehp.11449>
- ZIGLER, C. M. and DOMINICI, F. (2014). Point: Clarifying policy evidence with potential-outcomes thinking—beyond exposure-response estimation in air pollution epidemiology. *Am. J. Epidemiol.* **180** 1133–1140.