American Journal of Epidemiology, 2025, 00, 1–8
https://doi.org/10.1093/aje/kwaf252
Advance access publication date November 17, 2025
Practice of Epidemiology

Marginal structural models for quantifying the causal effects of exposure to ambient air pollution on progression of CT emphysema in the MESA lung and MESA air studies

Daniel Malinsky*[,1], Meng Wang[2], Rachel Heise[3], Carrie L. Pistenmaa[4,5], Eric A. Hoffman[6], Lianne Sheppard[7,8], Adam A. Szpiro[8], Andrew Laine[9], Elsa D. Angelini[9,10], Benjamin M. Smith[4,11], Joel D. Kaufman[12], R. Graham Barr[4]

[1]Department of Biostatistics, Columbia University, New York, NY, United States
[2]Department of Epidemiology and Environmental Health, University of Buffalo, Buffalo, NY, United States
[3]Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, United States
[4]Department of Medicine, Columbia University, New York, NY, United States
[5]Department of Medicine, Brigham and Women's Hospital, Boston, MA, United States
[6]Department of Radiology, University of Iowa, Iowa City, IA, United States
[7]Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA, United States
[8]Department of Biostatistics, University of Washington, Seattle, WA, United States
[9]Department of Biomedical Engineering, Columbia University, New York, NY, United States
[10]Telecom Paris, Institut Polytechnique de Paris, Palaiseau, France
[11]Department of Medicine, McGill University, Montreal, QC, Canada
[12]Department of Environmental and Occupational Health Sciences, Medicine, and Epidemiology, University of Washington, Seattle, WA, United States

*Corresponding author: Daniel Malinsky, R649 722 W. 168th Street, New York, NY 10032 (d.malinsky@columbia.edu)

Abstract

Associations between exposure to ambient air pollution and progression of emphysema have been identified in longitudinal observational studies. However, previous work has not used statistical causal inference methods tailored to address bias from time-varying confounding. The objective of this study is to propose an analytical approach for estimating longitudinal health effects of air pollution while accounting for time-varying confounding using marginal structural models and to re-analyze data on air pollution and emphysema progression from the Multi-Ethnic Study of Atherosclerosis using this analytical approach. We estimate weights for continuous exposure levels using two techniques: quantile binning of the exposure and a semiparametric model for the requisite conditional densities. The latter approach incorporates flexible machine learning methods. We find evidence for the harmful effects of ambient ozone pollution during study follow-up on the progression of emphysema, consistent with previously reported results. We find no evidence of effects of NO_x during study follow-up. This investigation demonstrates that analyses based on marginal structural models are feasible in studies of the health effects of air pollution and may address possible sources of bias that traditional regression-based methods fail to address. Further investigation is warranted to understand differences between our findings and previously published results.

Key words: air pollution; emphysema; causal inference.

Introduction

Chronic lower respiratory disease was the fourth leading cause of death in the United States and third leading cause of death worldwide in 2019. [1,2] Chronic obstructive pulmonary disease (COPD) and pulmonary emphysema are two of the largest contributors to these mortality figures. Ambient air pollution is an important risk factor for poor health generally and has been linked to both short-term and long-term respiratory morbidity.[3,4] Recent work documents associations between several key pollutants—ambient ozone (O_3), fine particulate matter ($PM_{2.5}$), oxides of nitrogen (NO_x), and black carbon measured at study baseline, as well as O_3 and NO_x measured over a median 10-year follow-up period, with the progression of emphysema assessed quantitatively by computed tomographic (CT) imaging and decline in lung function.[5] This evidence is derived from a longstanding longitudinal observational study: the Multi-Ethnic Study of Atherosclerosis (MESA). From the perspective of policy, a rigorous and robust understanding of the causal role of these air pollution exposures, typically based on multiple lines of evidence, has been identified as essential to informing air quality regulations, as evidenced by recent discussions among policymakers, governmental agency advisors, and academic researchers.[6-8] Eliminating

Received: February 17, 2025. **Accepted:** October 03, 2025

possible sources of bias is key to informing possible regulatory changes and emissions standards.

One source of potential bias that is particularly salient in causal inference from longitudinal observational studies that incorporate time-varying exposures, such as the analyses of air pollution concentrations over follow-up in MESA, is the existence of time-varying confounding. Time-varying confounders are measured covariates that change over the course of the study that are both predictive of (future) exposure and potentially affected by (past) exposure.[9] An example might be smoking behavior: smoking may be correlated with air pollution exposure at baseline, and later smoking may be causally affected by prior air pollution exposure over follow up (eg, if air pollution induces negative health symptoms such as difficulties breathing, leading a participant to quit or alter smoking behavior). Other possible time-varying confounders may measure health status (eg, biomarkers for cardiovascular health) or socioeconomic factors. In the context of multiple correlated time-varying pollutants, one pollutant may act as a time-varying confounder vis-à-vis another pollutant.

The challenge of time-varying confounders is that such variables are both "pre-exposure" (hence may act as confounders) and "post-exposure" (so may act as causal intermediates and also colliders on pathways to outcomes if there are unmeasured common causes), thus both naive adjustment and failure to adjust would induce bias for estimating the causal effect of interest.[9] Additional comments on bias from time-varying confounding may be found in Appendix S1 of the supplementary materials. Importantly, including time-varying confounders as adjustment variables in standard regression models is not sufficient to analytically control for this source of bias. Marginal structural models (MSMs) were introduced to support valid causal inference from longitudinal studies with time-varying confounding.[10-12] The threat of bias from time-varying confounding may be addressed by estimating MSM parameters with carefully constructed inverse probability weights (IPW) that sequentially adjust for time-varying covariates in a consistent way.

Another second potential benefit of analysis based on MSMs, as compared to typical regression-based approaches to estimating causal effects, is that the approach does not require correctly specifying a statistical model for the outcome as a function of covariates, ie, the approach remains agnostic to whether outcomes and covariates are linearly or non-linearly related. This may mitigate the threat of bias from model misspecification—the challenge in the regression-based approach lies in correctly specifying the form of the functional relationship between the outcome (eg, percent emphysema assessed by CT) and all potential confounders, which may be a list of a dozen or more variables. Often, linear (mixed) models are specified for analytical convenience, but even if the exposure-outcome relationship is linear, nonlinearities in the relationship between confounders and the outcome can also introduce bias into effect estimates.[13] The MSM approach does not require specifying a model for the outcome as a function of covariates, instead requiring specification of a series of models for the exposure. Though misspecification of these models may also introduce bias, the results may be robust insofar as the estimated weights produce the requisite "balance" across exposure groups.[14] At a minimum, the MSM approach is orthogonal to the regression-modeling approach and so can serve as a useful robustness check on the validity of results obtained by regression.

It has been challenging, however, to estimate MSMs with continuous exposures such as ambient air pollution levels.[15,16] Calculating the requisite inverse probability weights involves (repeated) conditional density estimation, which is difficult in practice. In this work, we reexamine the associations of 3 air pollution exposures (O_3, $PM_{2.5}$, and NO_x) that were assessed as time-varying exposures over follow-up with longitudinal progression of percent emphysema from MESA, taking an analytic approach based on MSMs. We propose 2 distinct estimation strategies to constructing inverse probability weights that adapt existing approaches that have so far been applied in point-exposure studies.[17-21] The first, simpler, strategy is to approximate the requisite conditional densities that appear in inverse probability weights by quantile binning: to compute probability weights, exposure concentrations are divided into some number of quantile "bins," and the probability that the observed exposure level falls within a given "bin" is estimated using a multinomial linear regression. It is straightforward to construct conditional density ratios from this procedure, and quantile binning has previously shown promising performance in a simulation study.[22] The second estimation strategy assumes a flexible semiparametric model for each conditional density that appears in the sequential weighting formula and uses an ensemble of machine learning methods to estimate components (mean and variance) of each density. Neither of these approaches to estimating inverse probability weights have been applied, to the best of our knowledge, to estimate the causal effects of individual-level air pollution exposures in longitudinal studies with repeated outcomes.

While an analytical approach based on marginal structural models cannot on its own eliminate all sources of potential bias in observational causal inference (eg, unmeasured confounding, measurement error), this approach is well-suited to address possible bias from improper handling of time-varying confounders in the context of time-varying exposures, as well as misspecification of the relationship between outcomes and covariates. Thus, the methodology presented here may be useful broadly in longitudinal studies that assess the effects of other (continuous) time-varying exposures and other health outcomes.

Methods

Sample

The MESA studied recruited participants from 6 US metropolitan areas (Baltimore, MD, Chicago, IL, Los Angeles, CA, New York, NY, St. Paul, MN, and Winston-Salem, NC) beginning in the year 2000.[23] Participants were aged 45-84 years and free of clinical cardiovascular disease at baseline. The study was designed to recruit participants who self-reported white, African American, Hispanic, or Asian race/ethnicity, defined by the 2000 US Census criteria, according to specified recruitment goals to create a more representative and diverse cohort than in prior studies. The MESA Air Pollution Study (MESA Air) recruited additional participants meeting the same enrollment criteria from 2005 to 2007.[24] The MESA Lung Study performed spirometry on a subsample of 3965 participants in 2004 to 2006, as well as additional participants in 2010-2012 and 2017-2018[25] and quantified emphysema on all MESA CT scans. The current analysis uses the same sample studied previously[5]; more detailed summary of the data and variable definitions may be found therein. The MESA and the ancillary studies were approved by institutional review boards at all sites and all participants provided informed consent.

Outcomes

Percent emphysema was calculated based on all usable CT scans, including cardiac and full-lung scans. Percent emphysema was defined as the percentage of lung pixels below −950 Hounsfield units (HU), with adjustment for attenuation of air outside the

chest to account for earlier scanner variation. Because cardiac CT images include the lower 66% of the lungs, approximately, the upper third of full-lung scans was excluded to follow the same lung region over time. Additional details related to lung imaging in MESA and outcome definition can be found in previous work.[5]

Exposures

The MESA Air study developed spatiotemporal exposure models for O_3, $PM_{2.5}$, and NO_x in each study region based on measurements (1999-2018) from the US Environmental Protection Agency (EPA) Air Quality System and spatially dense cohort-specific monitoring. These models have been described in detail previously.[24] Baseline exposure was assessed as mean pollutant concentration for the year of the baseline examination in 2000 and exposure over follow-up is assessed as time-varying mean concentration between the year of the baseline examination and follow-up time of the repeated outcome measure. Earlier work has separately evaluated the effects of exposure in a single time period (baseline exposure) and time-varying exposure (baseline and over follow-up).[5] We focus on the latter here since the challenge of time-varying confounding arises with time-varying exposure.

Covariates

Following earlier work,[5] covariates include baseline age, sex, race/ethnicity, study region, education, height, weight, body mass index (BMI), smoking (status, pack-years, cigarettes per day for current smokers, and secondhand smoke exposure), income, employment outside the home, physical activity, and neighborhood socioeconomic status index. Additionally, the data includes several variables related to CT scanner type, model, and scan parameters to capture variation and change in scanner technology. (In contrast to an earlier work,[5] we do not include long-term mean city-specific temperature, since mean city temperature is uninformative for emphysema conditional on remaining covariates, collinear with study region, and inclusion of temperature only introduces additional variability into effect estimates.)

We distinguish in our terminology here among three different types of covariates: time-varying confounders, other confounders, and precision variables. We use "time-varying confounders" to refer to covariates associated with outcomes that (1) change over time and (2) predict subsequent exposure but are also affected by earlier exposure, ie, they are intermediate variables along causal pathways from (earlier) exposure to outcomes. "Other confounders" (here all static covariates measured at baseline) are covariates that affect outcomes and subsequent exposures. These may be associated with later exposure due to "earlier," perhaps unmeasured, common determinants. More formal definitions/criteria exist,[10,26] but the key distinguishing feature germane to this longitudinal setting is whether the covariate may be affected by earlier exposure, since this feature is what makes traditional approaches to covariate adjustment invalid, as elaborated below. "Precision variables" are covariates that may be time-varying but are not associated with (neither causes nor effects of) exposure, only associated with outcomes. These may be straightforwardly incorporated as adjustment variables in outcome models to partially control for measurement error in the outcome.

In this analysis, we treat BMI and smoking behavior (status, cigarettes per day, and secondhand smoke exposure) as time-varying confounders. For each air pollutant, we include the remaining two pollutants as time-varying confounders in the "multipollutant" models reported below and additionally present "single" pollutant models that do not adjust for other pollutants. This reflects the possibility that for each target pollutant, the remaining two pollutants may act as time-varying confounders, though there is uncertainty about the causal "ordering" among pollutants. Computed tomography scanner variables are precision variables and included in all outcome models.

Analytical approach

The goal of our analysis is to quantify precisely the effect of hypothetical changes to air pollution exposure on progression of emphysema in US adults. Let Y_i denote the outcome, ie, percent emphysema assessed by CT, for individual i $(i = 1, \ldots, n)$, and let denote A_i denote their level of exposure to 1 specific pollutant of interest. Throughout, uppercase letters (A) will refer to random variables and lowercase (a) to specific values. The potential outcome (a.k.a. counterfactual) random variable $Y_i(a)$ denotes the outcome that would be observed if, possibly contrary-to-fact, individual i were "assigned" to exposure level $A_i = a$. The potential outcomes notation is an important ingredient of formal approaches to causal inference because it enables a distinction between values of variables that are observed in data (Y_i) versus would be observed under hypothetical (or counterfactual) changes to exposure level (from $Y_i(a')$ to $Y_i(a)$ for 2 levels of exposure a' to a) but not directly observed.

In this study, the exposure level varies continuously over a range of observed concentration values and the exposure, outcome, and covariates are measured on a repeated basis longitudinally at multiple visits over time. That is, the data consists in vectors $(X_i, L_{iv}, A_{iv}, Y_{iv})$ for multiple visits $v = 1, \ldots, m$ where X and L denote covariates. In the context of MESA, v corresponds to 1 of 6 exams (with $v = 1$denoting baseline). Some covariates X_i are measured at baseline and do not change over the course of the study (eg, demographics) and others L_{iv} are time-varying covariates. Exposures and outcomes are assessed at each visit, with A_{iv} denoting cumulative average air pollution concentration from baseline up to visit v.

See Figure 1 for a causal diagram representing the assumed longitudinal design, illustrated for 2 visits. We will assume, as encoded in this diagram, that observed baseline and time-varying confounders are sufficient to control for unmeasured confounding in this study. We denote the exposure history up to visit v by $\overline{A}_v = (A_1, \ldots, A_v)$ and likewise use the overbar notation as shorthand for history of other quantities $\left(\overline{L}_v, \overline{Y}_v\right)$. (Since exposure is assessed between year of baseline exam and follow-up exam, this justifies the inclusion of edges from A_v to Y_v.) Since the entire history of an individual's exposure is possibly relevant to their disease progression, the causal quantity of interest is $E\left[Y_{iv}\left(\overline{a}_v\right)\right]$, ie, the expected disease level for individual i at visit v under a hypothetical concentration history $\overline{A}_v = \overline{a}_v$. Marginal structural models are models for the concentration-response function $E\left[Y_{iv}\left(\overline{a}_v\right)\right] = g\left(\overline{a}_v; \beta^*\right)$, where g is some functional form parameterized by β^*. In the present analysis we specifically consider the following mixed linear MSM:

$$E\left[Y_{iv}\left(\overline{a}_v\right) \mid \delta_i^*, \tau_i^*, t_{iv}\right] = \delta_i^* + \tau_i^* t_{iv} + \beta_0^* + \beta_1^* t_{iv} + \beta_2^* a_{i1} + \beta_3^* t_{iv} a_{iv} \quad (1)$$

The MSM posits a linear dose–response relationship between cumulative air pollution and progression of emphysema. Here our target causal parameter is β_3^*, an interaction between cumulative air pollutant exposure and time (in years since baseline), which reflects the effect of a one-unit change in cumulative exposure during the period of observation on rate of change in

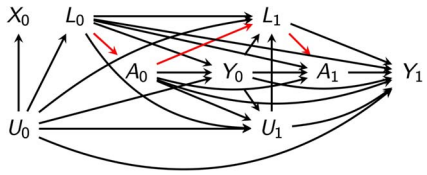

Figure 1. The DAG for a longitudinal study, shown for only 2 timepoints. Exposures are denoted by A, outcomes by Y, static/baseline confounders by X, ,and time-varying confounders by L. (Edges from baseline confounders X_0 to all other variables are omitted for clarity.) U denotes unmeasured factors, which are assumed not to affect exposures directly (the no unmeasured confounding assumption). Colored arrows highlight paths on which time-varying confounders are both "pre" and "post"-exposure.

percent emphysema. Individual-level random intercept and slope are included to reflect autocorrelation across repeated outcomes on the same individuals. (Since the MSM is linear and random effects are mean-zero, the target parameter would be unchanged if we marginalized over the random effects.) In addition, CT-scanner precision variables were included in the model to control for differences in outcome measurement; these are omitted in the eqn (1) above. A nonlinear MSM may also be of interest; however previous work using this sample[5] estimated a flexible concentration-response curve using regression splines that was consistent with linearity in the region of the data with most support.

It is worth contrasting briefly the approach based on MSMs to traditional regression modeling approaches for longitudinal data with repeated measures. Often, researchers posit a regression model for $E[Y_i \mid A_{iv}, L_{iv}, X_i]$, the conditional expectation of observed (not potential) outcomes given exposures and covariates. For example, a linear mixed model:

$E[Y_{iv} \mid A_{iv}, L_{iv}, X_i, \delta_i, \tau_i, t_{iv}] = \delta_i + \tau_i t_{iv} + \beta_0 + \beta_1 t_{iv} + \beta_2 A_{i1} + \beta_3 t_{iv} A_{iv} + \beta_4 L_{iv} + \beta_5 X_i$.

Importantly, even holding fixed a common functional form, the parameters of the MSM are not the same as the parameters of this kind of regression model. The MSM parameters support a causal interpretation (they represent the expected change in disease outcome under hypothetical interventions that change concentration history), whereas the regression model parameters do not necessarily support a causal interpretation except under specific assumptions or design conditions.[11,27] Put another way, estimates of β_3, even in the absence of model misspecification or unmeasured confounding, will be biased when treated as estimates of the causal β_3^* parameter. See Robins et al.[28] for a general discussion of the relationship between MSMs and traditional longitudinal models for observed outcomes.

Consider, for example, the possible role of smoking behavior as a time-varying confounder. At baseline, participants who live in higher-pollution areas are more likely to smoke, perhaps due to common socioeconomic determinants of smoking behavior and residential distribution. See the causal diagram in Figure 2. The baseline association between smoking and air pollution exposure is represented by the bidirected arrow between smoking and air pollution at baseline (bidirected arrows represent associations induced by common determinants). Over follow up, more participants in the higher-pollution areas may develop respiratory problems (asthma or symptoms such as coughing/phlegm, etc.) and consequently decide to quit or reduce smoking. This is represented by the directed arrow, indicating a possible causal effect, from baseline exposure to smoking at follow-up. Evidence has been reported linking both air pollution exposure to respiratory symptoms[29] and respiratory symptoms to changes in smoking behavior.[30]

We implemented two strategies for estimating the requisite inverse probability weights for continuous air pollution exposure levels: a quantile binning approach and a semiparametric ensemble learning approach. Due to space limitations, these are described in detail in the supplemental materials (Appendix S2). Briefly, the quantile binning estimator involves dividing the exposure level at each visit into 5 quantiles and predicting the probability that each individual's exposure falls into the relevant quantile using multinomial regression. The semiparametric estimator involves, for each density in the sequential weights formula, an amalgam of kernel density estimation with a weighted combination of several prediction models to estimate conditional means and variances (linear models, lasso regression, generalized additive models, random forests, extreme gradient boosting, and support vector machines). The estimated density ratios were truncated at the 1st and 99th percentiles at each visit to protect against the influence of extreme weights.[31]

Results

There were 7069 participants in the sample with percent emphysema measures and 6890 of these had assigned outdoor residential air pollution concentrations. Detailed descriptive characteristics of study participants in this sample were previously reported.[5]

Figure 3 shows correlations between values at baseline and at follow-up (at 1 exam, Exam 3) for several time-varying covariates and air pollution concentrations. Correlations between each pair of air pollutants were very strong. The strongest (excluding autocorrelations) was a correlation coefficient of −0.90 between NO_x and O_3. Correlations between baseline pollution concentrations and cigarettes per day at follow up were quite weak (strongest: −0.06 for NO_x) but somewhat stronger for BMI (strongest: −0.13 for $PM_{2.5}$).

Our primary results are summarized in Table 1. Estimates from the linear mixed model approach for time-varying exposures over follow-up in Wang et al.[5] were replicated and included in the third column for comparison (The linear mixed model estimates that serve as a baseline comparison to our approach differ slightly from the estimates reported previously with the same sample. The main reason for this is that we discovered some minor data quality issues for some of the exposure measurements [a small number of exposure values treated as "missing" even though they were not in fact missing in the data] so the exposure values were recalculated. A second reason is that we do not adjust for city-specific mean temperature, which should not be considered a confounder for the effect of interest in our setting [though city temperature is related to exposure levels, it is not informative for the outcome conditional on other covariates]. Finally, our analysis is performed in the R statistical software whereas earlier analysis employed SAS, which uses slightly different optimization routines for estimating mixed model parameters.). The effect of ozone on percent emphysema was estimated with a roughly consistent

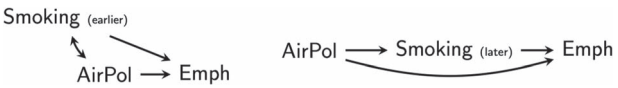

Figure 2. DAGs illustrating how smoking may be both associated with air pollution exposure at baseline and affected by air pollution over follow up. Abbreviations: AirPol, air pollution; Emph, emphysema.

Table 1. Estimated effects for air pollution exposures on progression of emphysema using three analytic approaches.

	Quantile binning	Semiparametric	LMM estimates
Ozone	0.073	**0.073**	**0.093**
(single)	CI: (−0.017, 0.162)	**CI: (0.006, 0.213)**	**CI: (0.003, 0.183)**
Ozone	0.053	0.064	**0.151**
(multipollutant)	CI: (−0.019, 0.150)	CI: (−0.006, 0.208)	**CI: (0.051, 0.250)**
NO_x	0.036	0.019	**0.117**
(single)	CI: (−0.029, 0.096)	CI: (−0.081, 0.105)	**CI: (0.040, 0.194)**
NO_x	0.014	0.020	**0.300**
(multipollutant)	CI: (−0.041, 0.103)	CI: (−0.114, 0.077)	**CI: (0.193, 0.404)**
$PM_{2.5}$	0.021	−0.153	−0.009
(single)	CI: (−0.113, 0.050)	CI: (−0.330, 0.063)	CI: (−0.115, 0.097)
$PM_{2.5}$	−0.045	0.004	−0.283
(multipollutant)	CI: (−0.139, 0.082)	CI: (−0.112, 0.044)	CI: (−0.430, −0.138)

Abbreviations: LMM, linear mixed regression model; NO_x, oxides of nitrogen; O_3, ozone; $PM_{2.5}$, particulate matter less than 2.5 mm in diameter.
Table shows effect of cumulative exposure to pollutant on longitudinal change in % emphysema over 10 years (per 3 ppb for O_3, per 10 ppb for NO_x, and per 2 mg/m^3 for $PM_{2.5}$). Multipollutant estimates adjust for the other 2 pollutants in estimated inverse probability weights. Cells in bold highlight estimates where 95% confidence intervals exclude zero. (95% CIs are calculated by bootstrap over 1000 replications for quantile binning and 200 replications for the semiparametric approach. Intervals for the LMM approach are model-based profile intervals.)

magnitude across all estimation strategies and models, with some variation in the size of accompanying 95% CIs (effect sizes ranged from 0.053% to 0.073% increase in emphysema per 3 ppb of O_3; lower CIs ranged from −0.019 to 0.006; upper CIs ranged from 0.150 to 0.213). The effect size estimated here was somewhat smaller in magnitude than the estimate produced via a linear mixed model approach, but nonetheless substantial and in the expected direction.

The effects of the other two pollutants, NO_x and $PM_{2.5}$, were small and indistinguishable from zero in all specifications. It is notable that effect estimates across the two estimation strategies, quantile binning and the semiparametric approach, are very similar to each other, with CIs from the semiparametric approach generally slightly wider compared to quantile binning. The present results for NO_x were inconsistent with the LMM results whereas those for $PM_{2.5}$ were consistent with the LMM results.

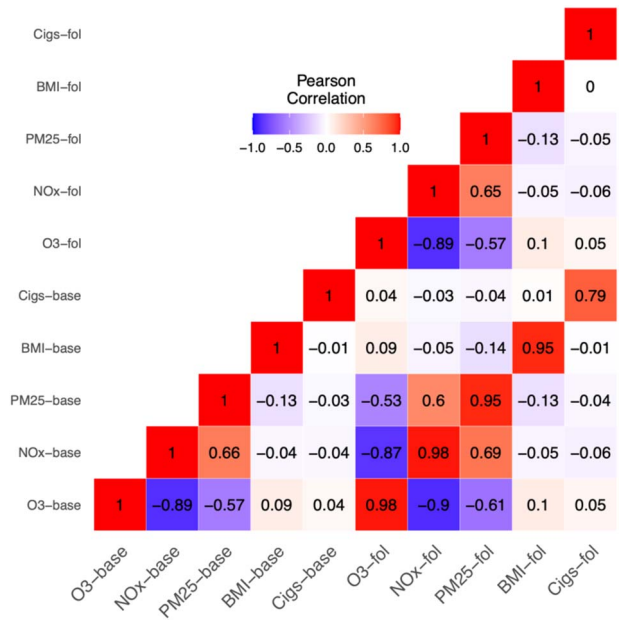

Figure 3. Pairwise correlations between exposures and some time-varying covariates: body mass index (BMI) and cigarettes-per-day (cigs) between baseline and follow-up (at Exam 3).

To probe the source of disagreement between our results and earlier results for NO_x, we examined alternative versions of our MSM approach that treated individual time-varying confounders as simple adjustment variables one at a time. Candidate time-varying confounders (ie, BMI and variables related to smoking) were removed from the weight model and instead included as adjustment variables in the outcome regression (the regression of percent emphysema on NO_x exposure, time since baseline, and scanner-related precision variables). This did not substantively alter the results or explain the difference between estimates for NO_x in the MSM vs LMM approach. Combined with the relatively weak correlations observed between (earlier) exposures and (later) time-varying confounders (Figure 3), this suggests that the difference between results is not likely due to bias from time-varying confounding. However, we did find that the estimates for NO_x were very sensitive to the handling of study region-by-time interaction. Following previous work, the LMM analysis specified an interaction between study region and time. When this interaction was removed from the model, regression estimates for the effect of NO_x were reduced by a factor of 10 and become non-significant. The MSM approach adjusts for region as a covariate in all the estimated weights but does not further condition on region (or region-by-time interaction) in the outcome model, which is a marginal model for the potential level of percent emphysema as a function of air pollution exposure only. Adding a region-by-time interaction in the MSM reproduced larger estimates for NO_x, consistent with the LMM results. One possible explanation of the difference vis-à-vis NO_x may be model misspecification: when two regression models that differ only in the specification of covariates (in this case, one including an interaction and the other not) produce substantively different estimates for the target effect, it is possible that at least one of the models is misspecified, though it is not possible to know which one. Another explanation is that the two analytical approaches, MSMs vs regression, are targeting different sets of parameters (as discussed above) and these will not necessarily coincide. Finally, the discrepancy between the 2 results for NO_x may also be the consequence of other violations of the requisite assumptions, including unmeasured confounding.

Discussion

We propose an analytical approach to estimating the effects of cumulative air pollution exposures on repeated outcomes using

marginal structural models. We introduce 2 estimation schemes for computing the requisite inverse probability weights, which allow for complex exposure distributions, without imposing any parametric model for the conditional density. A key advantage of marginal structural models is that they properly account for time-varying confounding, removing one possible source of bias that is often neglected in investigations of longitudinal time-varying air pollution effects. Furthermore, by treating "competing" exposures as time-varying covariates, MSMs may be well-suited to disambiguating the key drivers of health effects from among several possible correlated exposures.

Our analysis based on marginal structural models provides evidence for the harmful effects of ambient ozone pollution on chronic lung disease, specifically the progression of emphysema assessed quantitatively on CT scan. Estimators based on inverse probability weights are known to suffer from high variance (and hence lead to wide CIs), so our estimated CIs for the effect of ozone cover zero in some but not all specifications; point estimates are consistent across specifications and large in magnitude.

Previous studies have used MSMs with IPW to estimate the effects of air pollution exposures on other health outcomes, such as all-cause mortality[20] and hospitalization.[21] An important methodological distinction between these previous studies and ours is that they use a study design that is effectively a "point-exposure" rather than longitudinal repeated measures design, organizing observations into person-years (or person-seasons) and estimating a single propensity score for each observation to quantify how exposure in each person-time affects mortality or hospitalization in the following period. That is, these clustered designs only require estimating a single propensity score for each observation (person-time), whereas our longitudinal analysis, focusing on cumulative exposure over a long period with repeated outcomes on the same individual, requires estimating a product of (stabilized) propensity scores for each participant with potentially different covariates at each visit time, as detailed in our supplementary material (Appendix S2). Thus, the weight estimators introduced here enable researchers to use a flexible version of IPW to investigate cumulative exposure over a long period with repeated outcomes on the same individual.

An alternative analytical approach to the one presented here could be based on the parametric g-formula (a.k.a. g-computation). In this longitudinal setting, implementing the parametric g-formula would require correct specification of several "nuisance" models, including models for outcomes and time-varying covariates; if these models were correctly specified, the g-formula estimates would be unbiased and efficient.[32] However, the process of accurately specifying these several statistical models is challenging and error-prone. In realistic longitudinal scenarios where there is substantial risk of model misspecification, simulation studies indicate that which approach—MSMs with IPW versus parametric g-formula—performs better (eg, in terms of bias and CI coverage) depends on the details of the unknown data-generating process.[33,34] Indeed, Granger et al.[34] notes that g-formula, while generally more efficient than IPW, is more susceptible to bias due to model misspecification, and the risk of misspecification increases with an increase in the number of confounders. Although we do not implement the g-formula approach in this study, MSMs and other g-methods can complement each other as part of the wider triangulation of evidence on the effects of air pollution exposures.

The analysis here has important limitations. Practical considerations (discussed in the supplementary materials, Appendix S3) necessitated truncation of extreme weights and conditioning on a "reduced" covariate history in the weight estimation—both of these could lead to some bias. In our quantile binning estimator of the weights, exposure was binned into 5 quantiles—while varying this number slightly did not substantially affect results, with a much larger number of quantiles (eg, 10) the weights become unstable. In contrast, the semiparametric density estimation approach that uses an ensemble of machine learning methods does not require the user to specify some number of bins, but this estimator is very computationally intensive. Further, a causal interpretation of our analysis assumes that measured variables are sufficient to control for all sources of confounding and that measurement error is negligible, two assumptions that are not testable and are unlikely to be satisfied exactly.

Observational studies are key sources of evidence in environmental epidemiology because they can examine exposures at levels and scales that occur in the real world in the population of interest and follow participants over long periods of time to monitor disease progression. Data collected over these long time periods may be enlightening for exposures and diseases that unfold over the life course, but time-varying confounding introduces an analytical challenge when exposures are time-varying. Estimating effects based on marginal structural models while appropriately adjusting for time-varying confounders using sequential weighting can thus be a useful and flexible analytical approach in the study of environmental health effects in longitudinal settings.

Acknowledgments

The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org. This paper has been reviewed and approved by the MESA Publications and Presentations Committee. This publication was developed under the Science to Achieve Results (STAR) research assistance agreements, No. RD831697 (MESA Air) and RD-83830001 (MESA Air Next Stage), awarded by the U.S. Environmental Protection Agency (EPA). It has not been formally reviewed by the EPA. The views expressed in this document are solely those of the authors, and the EPA does not endorse any products or commercial services mentioned in this publication.

Supplementary material

Supplementary material is available at the *American Journal of Epidemiology* online.

Funding

D.M. was supported by the National Institutes of Health under award number K25ES034064 from NIEHS. This research was supported by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-,HC-95168, and N01-HC-95169 from the National Heart, Lung, and Blood Institute; and grants UL1-TR-000040, UL1-TR-0010,79, and UL1-TR-001420 from the National Center for Advancing Translational Sciences (NCATS). Funding support for the Lung CT dataset was provided by grants R01-HL077612 and RC1-HL100543.

Conflict of interest

Dr Pistenmaa discloses research funding to her institution from AstraZeneca and Sanofi and personal fees from InflaRx and Verona for advisory board participation. Dr Hoffman discloses that he is a founder and shareholder of VIDA Diagnostics, a company commercializing lung image analysis software developed, in part, at the University of Iowa. He is also an unpaid member of the Siemens Healthineers photon counting CT advisory committee. Dr Szpiro discloses funding from the Health Effects Institute. Dr Barr discloses funding from the American Lung Association and COPD Foundation outside the current work. All remaining authors declare that they have no conflicts of interest related to this work to disclose.

Data availability

The datasets supporting the conclusions of this article can be accessed by reasonable request to MESA Publication and Presentations Committee (http://www.mesa-nhlbi.org).

References

1. Murphy SL, Xu J, Kochanek KD, et al. Mortality in the united states, 2017. In: *NCHS Data Brief No 328*. National Center for Health Statistics. Published online 2018.
2. Roth GA, Abate D, Abate KH, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*. 2018;392(10159):1736-1788. https://doi.org/10.1016/S0140-6736(18)32203-7
3. Rice MB, Ljungman PL, Wilker EH, et al. Long-term exposure to traffic emissions and fine particulate matter and lung function decline in the Framingham heart study. *Am J Respir Crit Care Med*. 2015;191(6):656-664. https://doi.org/10.1164/rccm.201410-1875OC
4. Li J, Sun S, Tang R, et al. Major air pollutants and risk of COPD exacerbations: a systematic review and meta-analysis. *Int J Chron Obstruct Pulmon Dis*. 2016;11(1):3079-3091. https://doi.org/10.2147/COPD.S122282
5. Wang M, Aaron CP, Madrigano J, et al. Association between long-term exposure to ambient air pollution and change in quantitatively assessed emphysema and lung function. *JAMA*. 2019;322(6):546-556. https://doi.org/10.1001/jama.2019.10255
6. Dominici F, Zigler C. Best practices for gauging evidence of causality in air pollution epidemiology. *Am J Epidemiol*. 2017;186(12):1303-1309. https://doi.org/10.1093/aje/kwx307
7. Goldman GT, Dominici F. Don't abandon evidence and process on air pollution policy. *Science*. 2019;363(6434):1398-1400. https://doi.org/10.1126/science.aaw9460
8. Carone M, Dominici F, Sheppard L. In pursuit of evidence in air pollution epidemiology: the role of causally driven data science. *Epidemiology*. 2020;31(1):1-6. https://doi.org/10.1097/EDE.0000000000001090
9. Mansournia MA, Etminan M, Danaei G, et al. Handling time varying confounding in observational research. *BMJ*. 2017;359:j4587. https://doi.org/10.1136/bmj.j4587
10. Robins JM. Association, causation, and marginal structural models. *Synthese*. 1999;121(1-2):151-179. https://doi.org/10.1023/A:1005285815569
11. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran ME, Berry D, eds. *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. Springer-Verlag; 2000:95-133. https://doi.org/10.1007/978-1-4612-1284-3_2
12. Robins JM, Hernán MÁ, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550-560. https://doi.org/10.1097/00001648-200009000-00011
13. Vansteelandt S, Dukes O. Assumption-lean inference for generalised linear model parameters. *J R Stat Soc Ser B Stat Methodol*. 2022;84(3):657-685. https://doi.org/10.1111/rssb.12504
14. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015;34(28):3661-3679. https://doi.org/10.1002/sim.6607
15. Imai K, van Dyk DA. Causal inference with general treatment regimes: generalizing the propensity score. *J Am Stat Assoc*. 2004;99(467):854-866. https://doi.org/10.1198/016214504000001187
16. Zhao S, van Dyk DA, Imai K. Propensity score-based methods for causal inference in observational studies with non-binary treatments. *Stat Methods Med Res*. 2020;29(3):709-727. https://doi.org/10.1177/0962280219888745
17. Zhu Y, Coffman DL, Ghosh D. A boosting algorithm for estimating generalized propensity scores with continuous treatments. *J Causal Inference*. 2015;3(1):25-40. https://doi.org/10.1515/jci-2014-0022
18. Wu X, Braun D, Kioumourtzoglou MA, et al. Causal inference in the context of an error prone exposure: air pollution and mortality. *Ann Appl Stat*. 2019;13(1):520-547. https://doi.org/10.1214/18-AOAS1206
19. Wu X, Braun D, Schwartz J, et al. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. *Sci Adv*. 2020;6(29):eaba5692. https://doi.org/10.1126/sciadv.aba5692
20. Wei Y, Yazdi MD, Di Q, et al. Emulating causal dose-response relations between air pollutants and mortality in the Medicare population. *Environ Health*. 2021;20(1):1-10. https://doi.org/10.1186/s12940-021-00742-x
21. Nethery RC, Josey K, Gandhi P, et al. Air pollution and cardiovascular and thromboembolic events in older adults with high-risk conditions. *Am J Epidemiol*. 2023;192(8):1358-1370. https://doi.org/10.1093/aje/kwad089
22. Naimi AI, Moodie EE, Auger N, et al. Constructing inverse probability weights for continuous exposures: a comparison of methods. *Epidemiology*. 2014;25(2):292-299. https://doi.org/10.1097/EDE.0000000000000053
23. Bild DE, Bluemke DA, Burke GL, et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol*. 2002;156(9):871-881. https://doi.org/10.1093/aje/kwf113
24. Cohen MA, Adar SD, Allen RW, et al. Approach to estimating participant pollutant exposures in the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Environ Sci Technol*. 2009;43(13):4687-4693. https://doi.org/10.1021/es8030837
25. Barr RG, Ahmed FS, Carr JJ, et al. Subclinical atherosclerosis, airflow obstruction and emphysema: the MESA Lung Study. *Eur Respir J*. 2012;39(4):846-854. https://doi.org/10.1183/09031936.00165410
26. VanderWeele TJ, Shpitser I. On the definition of a confounder. *Ann Stat*. 2013;41(1):196-220. https://doi.org/10.1214/12-AOS1058
27. Hernán MA, Robins JM. *Causal Inference: What If*. CRC Press; 2020.
28. Robins JM, Greenland S, Hu FC. Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated

binary outcome. *J Am Stat Assoc*. 1999;94(447):687-700. https://doi.org/10.1080/01621459.1999.10474168
29. Su JG, Vuong V, Shahriary E, et al. Health effects of air pollution on respiratory symptoms: a longitudinal study using digital health sensors. *Environ Int*. 2024;189:108810. https://doi.org/10.1016/j.envint.2024.108810
30. Finocchio E, Olivieri M, Nguyen G, et al. Effects of respiratory disorders on smoking cessation and re-initiation in an Italian cohort study. *Int J Environ Res Public Health*. 2021;18(3):903. https://doi.org/10.3390/ijerph18030903
31. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168(6):656-664. https://doi.org/10.1093/aje/kwn164
32. Loh WW, Ren D, West SG. Parametric g-formula for testing time-varying causal effects: what it is, why it matters, and how to implement it in Lavaan. *Multivar Behav Res*. 2024;59(5):995-1018. https://doi.org/10.1080/00273171.2024.2354228
33. Souli Y, Trudel X, Diop A, et al. Longitudinal plasmode algorithms to evaluate statistical methods in realistic scenarios: an illustration applied to occupational epidemiology. *BMC Med Res Methodol*. 2023;23(1):242. https://doi.org/10.1186/s12874-023-02062-9
34. Granger E, Davies G, Keogh RH. Investigating the causal effects of multiple treatments using longitudinal data: a simulation study. *arXiv*. Preprint posted online May 2, 2024. https://doi.org/10.48550/arXiv.2405.01110

American Journal of Epidemiology, 2025, **00**, 1–8
https://doi.org/10.1093/aje/kwaf252
Practice of Epidemiology