Environmental Research 165 (2018) 228–234

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Assessing the cumulative health effect following short term exposure to multiple pollutants: An evaluation of methodological approaches using simulations and real data

Sophia Rodopoulou[a], Klea Katsouyanni[a,b], Pagona Lagiou[a], Evangelia Samoli[a,*]

[a] Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, 75 Mikras Asias Str, 115 27 Athens, Greece

[b] Department Population Health Sciences and Department of Analytical, Environmental and Forensic Sciences, School of Population Health & Environmental Sciences, King's College London, UK

ARTICLE INFO

Keywords:
Air pollution
Cumulative effect
Multi-pollutant models
LASSO
Exposure score

ABSTRACT

Background: Assessment of the cumulative effect of correlated exposures is an open methodological issue in environmental epidemiology. Most previous studies have applied regression models with interaction terms or dimension reduction methods. The combined effect of pollutants has been also evaluated through the use of exposure scores that incorporate weights based on the strength of the component-specific associations with health outcomes.

Methods: We compared three approaches addressing multi-pollutant exposures in epidemiological models: main effects models, the adaptive least absolute shrinkage and selection operator (LASSO) and a weighted exposure score. We assessed the performance of the methods by simulations under various scenarios for the pollutants' correlations. We further applied these methods to time series data from Athens, Greece in 2007–12 to investigate the combined effect of short-term exposure to six regulated pollutants on all-cause and respiratory mortality.

Results: The exposure score provided the least biased estimate under all correlation scenarios for both mortality outcomes. The adaptive LASSO performed well in the case of low and medium correlation between exposures while the main effect model resulted in severe bias. In the real data application, the cumulative effect estimate was similar between approaches for all-cause mortality ranging from 0.7% increase per interquartile range (IQR) (score) to 1.1% (main effects), while for respiratory mortality conclusions were contradictive and ranged from − 0.6% (adaptive LASSO) to 2.8% (score).

Conclusions: The use of a weighted exposure score to address cumulative effects of correlated metrics may perform well under different exposure correlation and variability in the health outcomes.

1. Introduction

The adverse effects of air pollution on health have been documented extensively. The Global Burden of Disease Study (Lim et al., 2012) classified air pollution (ambient particles) among the ten most important health risk factors worldwide. The vast majority of studies investigated the adverse effects related to exposure to individual air pollutants (gaseous or particulate) and assessed the sensitivity of their findings in two pollutant models. WHO (2013) The correlation between different pollutants often prohibits their simultaneous assessment. However, the air masses contain a mixture of pollutants at different concentration levels depending on their emission sources and the prevailing atmospheric conditions. Dominici et al. (2010) were among the first to highlight the importance of the investigation of health effects under the "one atmosphere" perspective. The REVIHAAP review (WHO, 2013) also emphasized the need to properly allocate the effects between different pollutants, urging future research to account for the simultaneous effects of multiple pollutants.

Assessing the impact of environmental exposures to human health in the context of the "one atmosphere" approach involves many challenges, including the control of possible synergies, the attribution of effects to individual pollutants taking into account collinearity, (Sun

Abbreviations: LASSO, Least Absolute Shrinkage and Selection Operator; AMEs, Additive Main Effects; ERFs, exposure-response functions; MSE, Mean Square Error; CI, Confidence Interval; IQR, Interquartile Range; $PM_{2.5}$, Particulate matter less than 2.5 μm in diameter; $PM_{2.5-10}$, Particulate matter between 2.5 and 10 μm in diameter; NO_2, Nitrogen dioxide; SO_2, Sulfur dioxide; CO, Carbon monoxide; O_3, Ozone; df, degrees of freedom

* Corresponding author.
E-mail address: esamoli@med.uoa.gr (E. Samoli).

https://doi.org/10.1016/j.envres.2018.04.021
Received 29 January 2018; Received in revised form 3 April 2018; Accepted 19 April 2018
0013-9351/

et al., 2013) but also measurement error impact on the health effect estimates. Bergen et al. (2016) The approach to solve these problems requires collaboration across disciplines (Dominici et al., 2010). Billionnet et al. (2012) reviewed appropriate statistical methods focusing on cross-sectional studies. They presented the Hierarchical Bayesian approach, dimension reduction methods, clustering, recursive partitioning, and logic regression as potential methods and concluded that researchers should base their choice on the specific characteristics of each data set.

Sun et al. (2013) used simulated and real data under cross-sectional and time series designs in order to assess the performance of five methods: the deletion/substitution/addition (DSA), the least absolute shrinkage and selection operator (LASSO), the Bayesian model averaging (BMA), the supervised principal component analysis (SPCA) and the partial least-squares regression (PLSR). They concluded that there is no optimal method across all scenarios, while the choice of the method should depend on the goal of the study. Among the methods reviewed, the LASSO seemed to work well under various study designs and parameter settings due to its robustness in estimation of regression coefficients and its increased power. Winquist et al. (2014) used epidemiological time series models with interaction terms between pollutants to assess their joint effect on pediatric asthma emergency department visits in Atlanta, US, and discussed that collinearity was not that high in relation to the sample size, since the effect estimate variances did not appear severely inflated.

Roberts (2006) proposed the introduction of pollutants' weights to epidemiological models and their estimation concurrently with models' parameters, while Park et al. (2014) proposed a two-step procedure: initial application of single and multi-pollutant models, identification of significant associations between exposures and the specific outcome and development of a weighted score with weights based on the estimated effects from models applied to a subset of the study data, and application of the score in epidemiological models using the rest of the data. Oakes et al. (2014) categorized approaches for multi-pollutant exposures into two broad groups, while Davalos et al. (2017) reviewed and classified statistical methods that investigate the association between short-term exposure to multi-pollutant mixtures and health effects in five broad groups: additive main effects (AMEs), effect measure modification, unsupervised and supervised dimension reduction and nonparametric methods. Although they advise to choose a statistical method based on the characteristics of the specific study, they highlighted the lack of studies that compare the performance of different statistical strategies in order to enhance knowledge for the development of optimal methods under varying circumstances.

In this study, we evaluate the performance of three indicative statistical approaches addressing the cumulative effect of multi-pollutant exposure: an additive (in terms of the model's linear predictor) main effects model, a refinement of LASSO named adaptive LASSO (Zou, 2006) as a dimension reduction method that uses penalized variable selection before the application of a main effects model and a weighted exposure score. We assess their performance using simulated time-series data under the context of a Poisson regression allowing for overdispersion, for the investigation of the effects of short-term exposure to six regulated air pollutants on mortality outcomes. We also apply the three approaches to real data as an example.

2. Materials and methods

2.1. Data

We collected data on daily concentrations of six air pollutants from the fixed monitoring sites of the Ministry of Environment & Energy in Athens, Greece for 2007–12: particulate matter with diameter less than 2.5 μm ($PM_{2.5}$, 24 h mean, $\mu g/m^3$) and 2.5–10 μm ($PM_{2.5-10}$, 24 h mean, $\mu g/m^3$), nitrogen dioxide (NO_2, 24 h mean, $\mu g/m^3$), sulfur dioxide (SO_2, 24 h mean, $\mu g/m^3$), carbon monoxide (CO, 8 h maximum, mg/m^3) and ozone (O_3, 8 h maximum, $\mu g/m^3$). We also collected data on daily number of deaths from all, excluding external, causes (International Classification of Disease, 10th Revision, ICD10: A00-R99) for ages > 15 years and for respiratory non-malignant causes (ICD 10: J00-J99) for ages 15–74 years. The mortality outcomes represent different numbers of observed events; all-cause mortality represents a large number of events per day, while respiratory non-malignant mortality has a small number of events and less variance compared to all-cause mortality. Data on daily mean temperature (°C) and mean relative humidity (%) were also collected.

2.2. Multi-pollutant statistical approaches

We assessed three different approaches in a Poisson regression allowing for overdispersion to investigate the effects of short-term exposure to six air pollutants on mortality outcomes. These approaches were selected as representative of commonly previously applied methods, in particular main effects models, dimension reduction approaches and composite indices based on air quality or prior factor analysis (Davalos et al., 2017). The general notation of the Poisson models was

$$\log \mathrm{E}[\mathrm{Y_t}] = \beta_0 + \sum\nolimits_{j} \beta_j \mathrm{P}_j^{t-1} + \text{ confounders}, \tag{1}$$

where Y_t the number of deaths at day t, P_j^{t-1} the concentration of pollutant j at day t-1 (lag 1), β_j the coefficient of pollutant j and β_0 the intercept, while the models were adjusted for possible time-varying confounders such as seasonality, long-term trends and meteorology; j ranges from one to six according to the approach used, as discussed below. We are interested in the estimation of the join effect of all six pollutants P^{t-1} on the health outcome Y_t.

Firstly, we included in the models all pollutants, i.e. in Eq. (1) j = 1…6, and assumed that each pollutant within the mixture had an additive effect in the linear predictor of the model. This method named AMEs (Davalos et al., 2017) is easy to construct and interpret, as it allows for the gradual inclusion of new terms. However, in the case of highly correlated variables, it involves the risk of collinearity and requires methods to correct inflated variances (Davalos et al., 2017). We applied this approach by estimating the joint effect as the exponential sum (across the six pollutants) of the product of each pollutant's coefficient with its interquartile range (IQR) based on Winquist et al. (2014) To control for potential multi-collinearity, the variance-covariance matrix of estimated coefficients was estimated using the sandwich estimator (Huber, 1967; White, 1980).

The second approach applied dimension reduction through pollutants selection based on the adaptive LASSO, i.e. in Eq. (1) the maximum value of j might range from one to six according to the adaptive LASSO selection. The LASSO, introduced by Tibshirani (1996) puts a ℓ1 penalty on the regression coefficients in a procedure that minimizes the sum of squared errors subject to the sum of the absolute values of the coefficients being less than a given value. (Dai et al., 2016a, 2016b) The LASSO estimates are defined as

$$\hat{\boldsymbol{\beta}}(\text{lasso}) = \operatorname{argmin}_{\boldsymbol{\beta}} \left\| \mathbf{y} - \sum\nolimits_{j=1}^{p} \mathbf{x}_j \beta_j \right\|^2 + \lambda \sum\nolimits_{j=1}^{p} |\beta_j|, \tag{2}$$

where λ is a nonnegative regularization parameter and p is the number of predictors in the model. The second term in Eq. (2) is the so-called "ℓ1 penalty". Zou (2006) The adaptive LASSO is a refinement of the LASSO that uses weights (w_j) for penalizing different coefficients in the ℓ1 penalty to achieve asymptotical normality and consistent selection. (Sun et al., 2013; Zou, 2006; Dai et al., 2016a, 2016b) This implies that the second term of the right part of Eq. (2) becomes $\lambda \sum_{j=1}^{p} w_j|\beta_j|$. The adaptive weight w_j is calculated as the inverse of the corresponding coefficient from a regression model ($w_j = 1/\beta_j^{\text{Poisson}}$). The use of LASSO for the study of adverse health effects of multiple pollutants was suggested by Roberts and Martin (2005). Using variables selection