Annals of Epidemiology xxx (2016) 1–9

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Review article

Current approaches used in epidemiologic studies to examine short-term multipollutant air pollution exposures

Angel D. Davalos MS [a], Thomas J. Luben PhD [b], Amy H. Herring ScD [a], Jason D. Sacks MPH [b,*]

[a] Department of Biostatistics, University of North Carolina, Chapel Hill
[b] National Center for Environmental Assessment, Office of Research and Development, US Environmental Protection Agency, Research Triangle Park, NC

A R T I C L E I N F O

Article history:
Received 8 August 2016
Accepted 27 November 2016
Available online xxx

Keywords:
Air pollution health effects
Joint effects
Multipollutant
Dimension reduction
Nonparametric methods
Interactions
Differential effects

A B S T R A C T

Purpose: Air pollution epidemiology traditionally focuses on the relationship between individual air pollutants and health outcomes (e.g., mortality). To account for potential copollutant confounding, individual pollutant associations are often estimated by adjusting or controlling for other pollutants in the mixture. Recently, the need to characterize the relationship between health outcomes and the larger multipollutant mixture has been emphasized in an attempt to better protect public health and inform more sustainable air quality management decisions.
Methods: New and innovative statistical methods to examine multipollutant exposures were identified through a broad literature search, with a specific focus on those statistical approaches currently used in epidemiologic studies of short-term exposures to criteria air pollutants (i.e., particulate matter, carbon monoxide, sulfur dioxide, nitrogen dioxide, and ozone).
Results: Five broad classes of statistical approaches were identified for examining associations between short-term multipollutant exposures and health outcomes, specifically additive main effects, effect measure modification, unsupervised dimension reduction, supervised dimension reduction, and nonparametric methods. These approaches are characterized including advantages and limitations in different epidemiologic scenarios.
Discussion: By highlighting the characteristics of various studies in which multipollutant statistical methods have been used, this review provides epidemiologists and biostatisticians with a resource to aid in the selection of the most optimal statistical method to use when examining multipollutant exposures.

Introduction

The results of epidemiologic studies that examine the association between individual air pollutants and health effects have contributed enormously to understanding how air pollution impacts health and the dramatic improvement in air quality that has occurred since the inception of the Clean Air Act. Although understanding the independent effects of exposure to a single pollutant is essential, scientists also recognize that under normal ambient conditions, humans are not exposed to individual pollutants in isolation but to a complex mixture of air pollutants. Recent publications convey this point by calling for research aimed at understanding the health effects of multipollutant exposures (i.e., the joint effect of two or more pollutants on a health outcome) with the aim of developing a catalog of statistical methods to support multipollutant analyses [1] that can inform the development of more sustainable air quality regulations [2,3].

Traditionally, epidemiologists examine whether there is evidence of an independent association between an individual pollutant on a health outcome (e.g., mortality) by including two or more air pollutants in a regression model and estimating the association attributable to each individual air pollutant after accounting for (or adjusting for) other measured pollutants co-occurring in the ambient air mixture. However, these types of models can become highly unstable when incorporating two or more pollutants that are highly correlated [2].

To examine the relationship between multipollutant exposures and health, new and innovative statistical methods are being developed and applied in epidemiologic studies. The purpose of this review is to highlight the variety of statistical methods currently available to examine the relationship between short-term exposures (i.e., single-day or multiday lags up to 1 week) to

The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

* Corresponding author. National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Mailcode B-243-01, Research Triangle Park, NC 27711. Tel.: +1-919-541-9729; fax: +1-919-541-2985.
E-mail address: sacks.jason@epa.gov (J.D. Sacks).

http://dx.doi.org/10.1016/j.annepidem.2016.11.016
1047-2797/Published by Elsevier Inc.

multipollutant mixtures and health effects. A number of these methods, specifically receptor modeling, have been used extensively to try and identify health risks associated with components and sources of fine particulate matter ($PM_{2.5}$), itself a multipollutant mixture. The multipollutant nature of $PM_{2.5}$ highlights a difficulty encountered when evaluating the current literature base of epidemiologic studies: the limited number of studies that focus specifically on examining the combined effect of multipollutant exposures to more than one criteria air pollutant (i.e., PM, O_3, NO_2, SO_2, and CO) on health. As such, for the purposes of this review, we focus on epidemiologic studies of multipollutant exposures that conduct a simultaneous evaluation of at least two criteria air pollutants, not studies focusing only on $PM_{2.5}$. Overall, this review is not intended to be a systematic evaluation of all available multipollutant statistical methods intended for use in short-term exposure epidemiologic studies but instead is meant to highlight the broad classes of statistical approaches available to epidemiologists and statisticians as they continue to design, conduct, and interpret multipollutant air pollution studies.

Methods

We conducted a broad literature search for studies including at least two criteria air pollutants (i.e., PM, O_3, NO_X, SO_X, CO). The broad literature search was a multistep process in which search strings were composed and then run through the PubMed and Web of Science databases. The search strings used for each pollutant are provided in Supplemental Table 1.

To the references retrieved by the broad literature search, a machine-learning algorithm was applied to segregate references into domains of epidemiologic or other (e.g., experimental) studies (see [4] for details). The algorithm, developed from a seed of known relevant references that focused on studies of air pollution and health, had recall greater than 90% but lower precision, meaning the bins contained some references not relevant for this review. As a result, a title screen was then performed to exclude nonrelevant references that were identified by the machine-learning algorithm. Finally, an abstract review was conducted to exclude any nonrelevant references that were not identified during the title screen. If we could not conclusively determine whether inclusion criteria were met from reviewing an abstract, we reviewed the reference's methods section. In addition, studies were identified for inclusion in several ways: specialized searches on specific topics, review of tables of contents for journals in which relevant papers may be published, identification of relevant literature by expert scientists, and review of citations in included studies. This is not intended to be a systematic review of the literature, but rather a broad overview of statistical methods and the feasibility and utility of their use to identify the combined effect of air pollutants in epidemiologic studies.

Results

Within this review, statistical methods are categorized according to the mixture effect assumptions (pollutant mixture relationship [PMR] specification) in the regression analysis. Based on the literature evaluated, five broad classes of statistical approaches were identified: additive main effects (AMEs), which are those methods that assume each pollutant within the mixture has an additive effect; effect measure modification (EMM), which are regression-based methods to examine whether the level of one or more pollutants modify the health effect associated with another pollutant or group of pollutants; unsupervised dimension reduction (UDR) that transforms multiple pollutants into a different set of variables independently of a health outcome of interest; supervised dimension reduction (SDR) where mixture transformation is dependent on the health outcome; and nonparametric methods, which are highly flexible methods that relax parametric assumptions of the interactive pollutant effects. Here, we use the language "effect" to refer to a general parameter of interest; we do not intend for the word "effect" to imply a necessarily causal association between exposure and outcome. The following sections provide a more detailed discussion of each broad class of multipollutant approaches along with the specific methods currently available.

Additive main effects

AME approaches, which consist of multipollutant or joint effects models with no multiplicative pollutant interaction terms, may be used to estimate joint associations of multiple air pollutants. The statistical methods within this category have appeal due to the intuitive construction of regression models, allowing for the straightforward inclusion of terms to examine the potential immediate, delayed, or prolonged association between air pollution and health through either single or multiday (e.g., distributed) lags.

Given the relative ease of construction and interpretability of AME models, surprisingly few air pollution studies use AME models to examine the combined association between multiple pollutants and health. Gold et al. [5] were one of the first to consider examining the combined effect of two pollutants ($PM_{2.5}$ and O_3) in a study of air pollution and lung function. They assessed pollutant-specific differences in the temporal relationship with lung function by including differing lag structures for each pollutant. Unlike Gold et al. [5], Schildcrout et al. [6] included the same lag structure for each pollutant (linear 3-day moving average) to examine the combined effect of a simultaneous increase in air pollutant concentrations on asthma exacerbations. The authors also decomposed the effect of ambient concentrations of pairs of pollutants (e.g., CO + NO_2, CO + PM_{10}) into a within- and between-subject component. Decomposing effects is a useful tool for revealing intraindividual and interindividual information and may be used for any of the other methods described in this review. However, the interpretability of effects and the additional number of coefficients to estimate will depend on the method chosen. Instead of focusing on two pollutant joint effects models, Winquist et al. [7] examined several pollutant mixtures (ranging from two to five pollutants) selected to represent pollutants that commonly occur together in ambient air or that might have common mechanisms leading to pediatric asthma emergency department (ED) visits. Collinearity was acknowledged as an issue in the pairwise CO + NO_2 model [6] and multipollutant models explored by Winquist et al. [7]. The AME specification does not in itself address multicollinearity and requires effect estimation procedures that can handle correlated variables to stabilize estimate precision, otherwise, estimates may not be obtainable or yield unreliable results.

Hierarchical models with AME specification have been used to study joint air pollution effects to overcome some difficulties with collinearity. Hierarchical models impose a distribution on effects (i.e., regression coefficients), where the effects can be assumed decomposed by a common property and pollutant-specific error resulting in pollutant effects being "shrunken" toward the effect of the common property with improved precision. When an AME specification is used within a hierarchical model, joint effects are immediately obtained on completion of the estimation procedure without need to aggregate pollutant-specific effects to obtain a joint effect. Suh et al. [8] demonstrated the use of such models by examining the joint impacts of 65 pollutants by nine chemical properties on the odds of daily cause-specific hospital admission through a two-stage hierarchical model (i.e., model is fit in a