Estimating the Health Effects of Exposure to Multi-Pollutant Mixture

CÉCILE BILLIONNET, MSc, DUANE SHERRILL, PhD,
AND ISABELLA ANNESI-MAESANO, PhD, MD, ON BEHALF OF THE GERIE STUDY

PURPOSE: Air pollution constitutes a major public health concern because of its ubiquity and of its potential health impact. Because individuals are exposed to many air pollutants at once that are highly correlated with each other, there is a need to consider the multi-pollutant exposure phenomenon. The characteristics of multiple pollutants that make statistical analysis of health-related effects of air pollution complex include the high correlation between pollutants prevents the use of standard statistical methods, the potential existence of interaction between pollutants, the common measurement errors, the importance of the number of pollutants to consider, and the potential nonlinear relationship between exposure and health.

METHODS: We made a review of statistical methods either used in the literature to study the effect of multiple pollutants or identified as potentially applicable to this problem. We reported the results of investigations that applied such methods.

RESULTS: Eighteen publications have investigated the multi-pollutant effects, 5 on indoor pollution, 10 on outdoor pollution, and 3 on statistical methodology with application on outdoor pollution. Some other publications have only addressed statistical methodology.

CONCLUSIONS: The use of Hierarchical Bayesian approach, dimension reduction methods, clustering, recursive partitioning, and logic regression are some potential methods described. Methods that provide figures for risk assessments should be put forward in public health decisions.
Ann Epidemiol 2012;22:126–141.

KEY WORDS: Air Pollution, Collinearity, Combined Effect, Correlated Pollutants, Multiple Pollutants.

INTRODUCTION

Air pollution constitutes a major public health concern because of its ubiquity and of its potential short-term and/or long-term health impact. Since the London smog episode, several publications have put forward the deleterious effects of many indoor and outdoor air pollutants on individual health of children and adults by considering each at one (single-pollutant approach). Most epidemiologic studies on health effects of air pollution have focused on understanding the effects of criteria air pollutants, ozone (O_3), nitrogen dioxide (NO_x), sulfur dioxide (SO_2), lead, carbon monoxide (CO), and particulate matter (PM), with each considered as having an isolated effect.

From the INSERM, UMR S 707, EPAR, F-75012, Paris, France (C.B., I.A.-M.); UPMC Univ Paris 6, UMR S 707, EPAR, F-75012, Paris, France (C.B., I.A.-M.); and Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ (D.S.).

Address correspondence to: Cécile Billionnet, MSc, UMR S 707 (EPAR), INSERM & UPMC, Department of Epidemiology of Allergic and Respiratory Diseases (EPAR), Paris, France. Tel.: +33 144738437. E-mail: billionnet@u707.jussieu.fr.

This work has been supported by Contract n 2006343 of European DGSANCO.

Received March 25, 2011. Accepted November 15, 2011.

Indeed, the real world is different: polluted air contains a complex mixture of particle and gas phase pollutants so that individuals are exposed to many air pollutants at once. Most pollutants are highly correlated to each other, so that an additive or synergic effect cannot be excluded. Consequently, single pollutant models may be difficult to interpret (1). In addition, individuals are exposed to differing amounts of air pollutants depending on the type of emissions. This has led to the need to consider exposure to many air pollutants at once and to develop an appropriate methodology adjusting for these correlations (multi-pollutant approach). This premised has been underlying in 2007 by the Scientific Committee "Health and Environmental Risks" of the European Commission (http://ec.europa.eu/health) that strongly encourages the evaluation of the combined effects of indoor air pollutants.

The development of methods estimating the adverse health effects of multiple exposures is thus an important topic to explore. Dominici et al. (2), Mauderly et al. (3) and Vedal and Kaufman. (4) published recently very informative papers in which they address the complexities of multi-pollutant health effects and related methods. Various difficulties exist when multiple exposure effects are assessed; in this paper, we concentrate on only statistical difficulties.

360 Park Avenue South, New York, NY 10010

1047-2797/$ - see front matter
doi:10.1016/j.annepidem.2011.11.004

Selected Abbreviations and Acronyms

PM = particulate matter
VOC = volatile organic compounds
OR = odds ratio
VIF = variance inflation factor
DSA = deletion/substitution/addition
BMA = Bayesian model averaging
RP = recursive partitioning
PCA = principal component analysis
SPCA = supervised principal component analysis
PLS = partial least-square
GAM = generalized additive model
PMF = positive matrix factorization

In most of the papers authors have analyzed multiple exposure effects either by fitting a regression with all exposures in the model or use a step-by-step algorithm, such as forward regression or stepwise regression, to produce a reduced model. However, these statistical standard methods, which simultaneously include multiple exposures in a single model but consider their impact independently, can lead to interpretation and estimation errors. That is why, over recent years, multiple pollutants are being investigated as a mixture effect with more complex statistical methods (5–7).

In this paper we address major approaches that can be used to investigate the multi-pollutant issue. After having highlighted the complexities of multi-pollutant assessment that prevent use of standard methods, we identify statistical methods that have been applied in the literature so far. Examples of investigations having applied such methods are reported, which allows assessing the associations between multiple pollutants and health outcomes. We also review other methods that could be adapted for such a study. For each statistical method presented, we examine the cons and pros, as well for the quality of the obtain results related to their ease of interpretation. Note that we focus our paper on methods adaptable for cross-sectional study, under the hypothesis that the concentrations of pollutants are measured at time t without any follow-up and targeted air pollution effects.

CONTEXT

High correlation may exist between pollutants. Outdoor traffic-related air pollutants such as NO_2 and PM and indoor volatile organic compounds (VOCs) are strongly correlated each other. McConnell et al. (8) found, in a study from California, correlations of 0.83 and 0.73 for NO_2 with $PM_{2.5}$ and PM_{10} respectively. In a survey conducted by the French Indoor Air Quality Observatory (9) in which we are investigating health effects of indoor air pollutants (10), Spearman's correlations for BTEX (i.e., benzene, toluene, ethylbenzene, and xylenes) concentrations in dwellings ranged from 0.56 to 0.96.

The concentration of one VOC is often related to the concentration of the others, with each household component emitting a given set of VOCs. In human beings, exposure to VOCs may result in a spectrum of illnesses ranging from mild, such as irritation, to very severe, such as cancer (11, 12). Health effects of outdoors air pollutants are also well established and include both short-term and long-term effects on morbidity and mortality, overall in the case of cardiorespiratory diseases (13). These effects have been seen at very low levels of exposure in many epidemiological studies. However, studies having identified health effects of air pollutants have considered each compound individually or have simply adjusted in the regression models for other pollutants.

Collinearity, where air pollutants are so highly correlated that it is impossible to come up with reliable estimates of their individual regression coefficients, poses a real problem if the purpose of the study is to estimate the contributions of individual predictors. In our Six Cities Study, the odds ratio (OR) between each health indicators and each air pollutants was obtained with the logistic regression model after adjustment for confounders and the highest correlated air pollutant (14). As a consequence of correlation computations, two 2-pollutants models were applied, including NO_2 and ozone (O_3) for one and SO_2 and PM_{10} for the other. The study of each VOC independently can lead to wrong conclusions, for example, highlighting an association between a specific VOC and a respiratory pathology actually caused by a second VOC very correlated with the first one.

That is why, in studying the health effects of VOCs, one needs a multivariate model. However, multicollinearity prevents one from using standard methods with several VOC concentrations as predictors in a single regression model. This problem exists for other air pollutants found indoors and outdoors. In addition, interaction between pollutants can occur, and more specifically, synergy. Synergy is defined as occurring if the effect of the combined exposure is greater than the sum of the effects of the two or more individual pollutants of the mixture (15). However, this method holds only when pollutant effects are linear. Considering a "sham combination" of two concentrations of the same pollutant: only pollutants with linear concentration–responses curve would be considered non-interactive, based on the sum of the effect (16). Instead of adding effects, we may add concentrations proportional to their effects. If the curve connecting concentrations of equal effect (isoboles) is linear, the mixture is not synergistic; if the curve is concave, the mixture is synergistic. However, there are several authors who remain skeptical about the general validity of this method when the log dose-effect curves of the individual agents are not parallel (17). Interdependence