Annals of Epidemiology 23 (2013) 198–203

Contents lists available at SciVerse ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

A Bayesian approach to modeling the interaction between air pollution and temperature

Borek Puza PhD, Steven Roberts PhD *

Australian National University, Canberra

ARTICLE INFO

Article history:
Received 18 September 2012
Accepted 12 January 2013
Available online 20 February 2013

Keywords:
Air pollution
Bayesian model
Cutpoint estimation
MCMC methods

ABSTRACT

Purpose: Investigating the interaction between particulate matter air pollution (PM) and temperature is important for quantifying the effects of PM on mortality. One approach is stratification—estimating the effect of PM within different temperature strata—but this treats the cutpoints that define the strata as fixed, when in fact they are unknown. The purpose of this paper is to propose a new approach that appropriately accounts for uncertainty regarding the cutpoints, and to apply this approach to data from two Australian cities.

Methods: We propose a Bayesian model which allows the effects of PM to differ within different temperature strata. The cutpoints that define the strata are parameters that are jointly estimated along with the other model parameters. This is in contrast with the standard stratification approach, where cutpoints are specified a priori and treated as fixed constants. Also, the Bayesian model is formulated in a way that ensures continuity in the effects of PM at the stratum boundaries. Markov chain Monte Carlo methods are used to perform the inferences.

Results: Analysis of daily data over several years provides evidence for an interactive effect between PM and temperature in Sydney and no support for such an effect in Melbourne.

Conclusions: The proposed Bayesian model provides a means for investigating interactions between PM and temperature which appropriately incorporates uncertainty.

Introduction

Over the last 10 years, there has been vigorous research on the health effects of both ambient particular matter air pollution (PM) and temperature [1–4]. The general consensus from this research is that PM and temperature are both associated with an increase in adverse health outcomes. In addition to these two lines of independent research, there have been studies investigating whether there are interactions between PM and temperature or whether temperature modifies the health effects of PM [5–7]. Essentially, these studies attempt to address the important question of whether the health effects of PM differ depending on the temperature. Knowing whether such interactions exist is clearly important from a public health and regulatory standpoint, in terms of quantifying the risk of exposure to PM and also providing appropriate advice to the general public. On a related note to the question of interactions between PM and temperature, there has also been recent interest in whether the effect of air pollution varies with season [8–10]. For example, Park et al. [9] find evidence that both season and temperature modify the mortality effect of air pollution.

Most studies that have investigated the interaction between PM and temperature use one or more of the following approaches: Product terms, stratification, and response surfaces [6,7]. The *product terms approach* involves fitting a model that includes, in addition to PM and temperature as main effects, a term equal to PM multiplied by temperature, allowing the effect of PM to differ depending on the temperature. This approach corresponds to the 'standard' method of allowing for interactions in statistical modeling. The *stratification approach* involves defining temperature strata and then fitting models that allow the effect of PM to differ between these. The *response surface approach* involves modeling the relationship between PM and temperature as a continuous function (or surface) of both variables. Each of these approaches has its advantages and disadvantages. In particular, the stratification approach is the most easily interpretable, but requires the specification of temperature cutpoints that determine the temperature strata. On the other hand, the product terms and response surface approaches do not require temperature cutpoints to be specified, but can be harder to interpret, particularly if the results are used to explicitly quantify the effect of PM. Additionally, with the response surface approach it can be difficult to allow for delayed effects of

* Corresponding author. Research School of Finance, Actuarial Studies and Applied Statistics, College of Business and Economics, Australian National University, ACT 0200 Australia. Tel.: +61 2 612 53470; Fax: +61 2 612 50087.
E-mail address: steven.roberts@anu.edu.au (S. Roberts).

1047-2797/$ – see front matter
http://dx.doi.org/10.1016/j.annepidem.2013.01.002

pollution and/or temperature. One suggestion is to use a response surface analysis to help inform the location of the temperature cutpoints for use in a stratification analysis [7]. The findings of studies that have investigated possible interactions between PM and temperature have been mixed in terms of whether or not significant interactions have been found [5].

The purpose of our study is to introduce a Bayesian model that allows for a possible interaction between PM and temperature. In this model we allow the effect of PM to differ depending on three temperature strata. Importantly, however, the temperature cutpoints that define the temperature strata are parameters in the model that are jointly estimated along with the effects of PM. This is in contrast with the stratification approach, where temperature cutpoints are specified a priori and then treated as fixed constants in the modeling process. A downside of the stratification approach is that the final modeling results treat the cutpoints as fixed and therefore may not adequately reflect the true statistical uncertainties associated with first investigating a range of potential cutpoints. This is avoided in our Bayesian model, where the cutpoints are included in the model and uncertainty about their location is explicitly incorporated in any subsequent analysis. This is an important step that ensures that any finding of a significant interaction between PM and mortality is not simply an artifact of multiple testing. We illustrate that our Bayesian method can detect interactions when they exist and apply the method to data from two Australian cities.

Materials

In this investigation, we use data from Australia's two largest cities, Melbourne and Sydney. For Melbourne we use data for the period May 19, 1995, to November 30, 2007, and for Sydney, April 3, 1993 to November 30, 2007.

All-cause mortality data (excluding external causes) were obtained from the Australian Bureau of Statistics for the Melbourne and Sydney statistical divisions. Measures of the daily temperature and dew-point temperature (both 24-hour averages) were obtained from data supplied by the Australian Bureau of Meteorology. PM of less than 10 μm (PM_{10}) is the measure of PM that will be used. For Melbourne, these data were obtained from the Environment Protection Authority Victoria and for Sydney from the Department of Environment, Climate Change and Water (DECCW NSW). The 24-hour average PM_{10} concentrations used in our study were obtained by averaging the values available from a sequence of monitors within each city. We choose a 24-hour average of PM_{10} because this is the most common measure of PM_{10} used in the literature [11]. However, we note that other possibilities exist, including a weighted average of hourly PM_{10} concentrations (possibly giving higher weight to higher concentrations) or the maximum hourly concentration. The utility of these other possible PM_{10} measures would be an interesting area for further research. Finally, for the purposes of our investigation missing values were imputed by taking an average of the values on the day before and day after (for each of the variables).

Methods

In the models fitted below the PM_{10} exposure measure that was used, denoted x_t, is the average of the current and previous 2 days' PM_{10} concentrations. The exposure measure of temperature that is used, denoted z_t, is the current day's average temperature. The use of the current day's temperature is a simplification that is used to enable us to focus on the development of our proposed Bayesian model. However, we stress that this does not mean that the effect of temperature on mortality lasts for a single day. Indeed, in the literature there is evidence that the effect of temperature on mortality can last for weeks [12,13]. In addition to the effects of temperature, the models fitted below also adjust for the effects of seasonal confounding, dew point temperature, and day of the week.

The standard stratification model

A standard stratification investigation of the interaction between temperature and PM_{10} proceeds by fitting the model:

$$\begin{aligned} &y_t \sim \text{independent Poisson}(\mu_t), \quad t = 1,\ldots,n \\ &\log(\mu_t) = \beta_1 + \beta_2 x_t + \beta_3 x_t I(z_t < L) + \beta_4 x_t I(z_t > H) + m_t^T \gamma \\ &m_t^T = (s(t, 4\text{ df/year}), s(z_t, 6\text{ df}), s(\text{dew}_t, 3\text{ df}), \text{dow}_t) \end{aligned} \tag{1.1}$$

where t indicates day, y_t is the mortality count (on day t), μ_t is the mean mortality count, x_t is the standardized 3-day average PM_{10} concentration, z_t is the standardized lag-0 temperature (in degrees Celsius on the original scale), $I(z_t < L)$ is an indicator variable taking the value 1 if temperature is below a fixed cutpoint L, $I(z_t > H)$ is an indicator variable taking the value 1 if temperature is above a fixed cutpoint H, and $m_t = (m_{1t},\ldots,m_{pt})^T$ represents a vector of potential confounding covariates. These confounders represent the underlying effects of time ($s(t$, 4 df/year)), temperature ($s(z_t$, 6 df)), dew point temperature ($s(dew_t$, 3 df)), and day of the week (dow_t). The current day's temperature is modeled as a natural spline with 6 degrees of freedom, the current day's dew point temperature is modeled as a natural spline with 3 degrees of freedom, day of the week is modeled as a categorical variable, and to control for seasonal confounding the effects of time are modeled as a natural cubic spline with 4 degrees of freedom per year.

In Model (1.1) the parameters representing the effects of air pollution are contained in the vector $\beta = (\beta_1,\ldots,\beta_4)^T$ (the first element of which is an intercept term), and the parameters representing the effects of the confounders are given by the vector $\gamma = (\gamma_1,\ldots,\gamma_p)^T$. Models similar to this have been used in a number of previous studies [7,14–16].

Note that β_2 represents the incremental effect of pollution between temperatures L and H, β_3 represents the additional incremental effect of pollution for temperatures below L, and β_4 represents the additional incremental effect of pollution for temperatures above H. Thus, the incremental effect of PM_{10} is equal to $\beta_2 + \beta_3$ if the temperature is less than L, β_2 if temperature is between L and H, and $\beta_2 + \beta_4$ if temperature is greater than H. It is clear that model (1.1) allows the effect of PM_{10} to differ depending on which of the three strata temperature falls in.

As mentioned, a downside of this approach is the fact that the cutpoints L and H are treated as fixed. This means that the model does not allow for the fact that a search over possible combinations of L and H may have been conducted before fitting the final model. Another potentially undesirable property of model (1.1) is the discrete nature (or "jump") in the effect of PM_{10} that occurs at each of the temperature cutpoints. The Bayesian model that we introduce includes L and H as parameters in the model and also allows for a smooth change in the effect of PM_{10} at each of the temperature cutpoints.

A Bayesian model

The Bayesian approach provides a convenient framework for dealing with the type of inferential problems described. Within the classical (or frequentist) paradigm, estimation of the cutpoints L and H is problematic and necessitates one or more ad hoc procedures, such as searching for the values of these cutpoints which minimize the Akaike information criterion. One downside of such