Science of the Total Environment 409 (2010) 300–306

Contents lists available at ScienceDirect

Science of the Total Environment

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

The short-term effect of air pollution on cardiovascular mortality in Tianjin, China: Comparison of time series and case–crossover analyses

Yuming Guo [a,*], Adrian G Barnett [a], Yanshen Zhang [b], Shilu Tong [a], Weiwei Yu [a], Xiaochuan Pan [c,*]

[a] School of Public Health, Institute of Health and Biomedical Innovation, Queensland University of Technology, Kelvin Grove, Brisbane, Queensland 4059, Australia
[b] Department of Environmental Pollution and Health, Chinese Research Academy of Environmental Sciences, Beijing 100012, China
[c] Department of Occupational and Environmental Health, School of Public Health, Peking University, Beijing 100191, China

A R T I C L E I N F O

Article history:
Received 5 August 2010
Received in revised form 6 October 2010
Accepted 11 October 2010
Available online 4 November 2010

Keywords:
Air pollution
Cardiovascular mortality
Case–crossover
Time series

A B S T R A C T

Background: Many studies have illustrated that ambient air pollution negatively impacts on health. However, little evidence is available for the effects of air pollution on cardiovascular mortality (CVM) in Tianjin, China. Also, no study has examined which strata length for the time-stratified case–crossover analysis gives estimates that most closely match the estimates from time series analysis.
Objectives: The purpose of this study was to estimate the effects of air pollutants on CVM in Tianjin, China, and compare time-stratified case–crossover and time series analyses.
Method: A time-stratified case–crossover and generalized additive model (time series) were applied to examine the impact of air pollution on CVM from 2005 to 2007. Four time-stratified case–crossover analyses were used by varying the stratum length (Calendar month, 28, 21 or 14 days). Jackknifing was used to compare the methods. Residual analysis was used to check whether the models fitted well.
Results: Both case–crossover and time series analyses show that air pollutants (PM_{10}, SO_2 and NO_2) were positively associated with CVM. The estimates from the time-stratified case–crossover varied greatly with changing strata length. The estimates from the time series analyses varied slightly with changing degrees of freedom per year for time. The residuals from the time series analyses had less autocorrelation than those from the case–crossover analyses indicating a better fit.
Conclusion: Air pollution was associated with an increased risk of CVM in Tianjin, China. Time series analyses performed better than the time-stratified case–crossover analyses in terms of residual checking.

1. Introduction

Many epidemiological studies have shown that short-term increases in ambient air pollutants are associated with an acute rise in mortality (Touloumi et al., 2005), hospital admissions (Dominici et al., 2006), and emergency hospital visits (Guo et al., 2009). However, most studies were conducted in western countries (where people have different demographic characteristics compared with Asian countries) and used time series and case–crossover analyses separately. Also, studies of the health effects of air pollution on mortality and morbidity in China, have mostly been conducted in Beijing (Guo et al., 2009; Guo et al., 2010), Shanghai (Kan et al., 2008), Shenyang (Xu et al., 2000), Wuhan (Qian et al., 2007), and Taiyuan (Zhang et al., 2007). To our knowledge, few studies have assessed the relationship between air pollution and cardiovascular mortality (CVM) in Tianjin, a large industrial city in northeastern China.

Time series and case–crossover analyses are the most common methods used to estimate the short-term effects of air pollution on health (Fung et al., 2003; Schwartz, 2004). Time series analysis allows for over-dispersion associated with the Poisson distribution and controls for long-term trend and seasonality using nonparametric or parametric splines. The generalized additive model (GAM) is often used to examine associations between air pollution and health (Dominici et al., 2002; Samet et al., 2000).

The case–crossover compares exposure during a case day when events occurred (e.g., deaths) with exposures in nearby control days to examine whether the events are associated with a particular exposure. Because the control days are selected close to the case days, seasonality is controlled for by design. Confounding for day of the week can be controlled by choosing control days with the same day of the week as the case days. Confounders related to individual characteristics (e.g., age, sex and smoking) are also controlled by design. There are many different designs for choosing control days

* Corresponding authors. Guo is to be contacted at School of Public Health, Institute of Health and Biomedical Innovation, Queensland University of Technology, Kelvin Grove, Brisbane, Queensland 4059, Australia. Tel.: +61 7 31383996; fax: 61 7 31383130. Pan, Room 706, Department of Occupational and Environmental Health, Peking University School of Public Health, No 38, Xueyuan Road, Haidian District, Beijing 100191, China; Tel./fax: +86 10 82802530.
E-mail addresses: guoyuming@yahoo.cn (Y. Guo), xcpan@hsc.pku.edu.cn (X. Pan).

0048-9697/$ – see front matter
doi:10.1016/j.scitotenv.2010.10.013

relative to a case day. We considered three: unidirectional, bidirectional and time-stratified.

A unidirectional design selects fixed control day(s) per case day only before or after the case day. For example, the air pollution exposure was selected seven days before or after the case day. This design does not control for trends over time in air pollution or health outcomes, and so is subject to bias (Greenland, 1996).

Bidirectional designs include the full-stratum bidirectional (Navidi, 1998), symmetric bidirectional (Bateson and Schwartz, 1999), and semi-symmetric case–crossover (Navidi and Weinhandl, 2002). The full-stratum bidirectional case–crossover was designed to include all exposures in the time series before and after the case days as control days. This design controls for time trends in exposure, but does not control for seasonal patterns in exposure or health outcomes (Bateson and Schwartz, 1999).

The symmetric bidirectional design uses control days both before and after the case day. This method can successfully control for seasonality in exposures and outcomes. However, there is the potential for selection bias, because the cases at the beginning or end of the data series have fewer control days for matching. Navidi and Weinhandl (2002) noted that the symmetric case–crossover design might still be biased by time trends from exposure.

The semi-symmetric design randomly selects a control day before or after the case day. This design can also control for long-term trends and seasonality. However, because only one control day is selected at a fixed interval, the estimates may still be biased (Levy et al., 2001). Lumley and Levy (2000) illustrated how selection biases do not appear when cases may occur at any time in the strata from which the controls are selected. This design is the initial principle of the time-stratified case–crossover. They demonstrated that most of the above designs are biased because the controls are not chosen independently of the case day. This bias is called the "overlap bias" and occurs in case–crossover designs with non-disjointed strata (Lumley and Levy, 2000). Janes et al. (2005) demonstrated that the overlap bias is not an issue for the time-stratified design.

As the time-stratified case–crossover uses fixed and disjointed time strata (e.g., calendar month), the overlap bias is avoided. Studies have shown that case–crossover analyses are equivalent to time series analyses (Basu et al., 2005; Fung et al., 2003). However, no study has examined which strata length gives results most similar to a time series analysis, and no study has used residual analyses to check the adequacy of time-stratified case–crossover models.

The aims of this study were to explore whether there was any short-term effect of air pollution on CVM in Tianjin, China, and to compare the time series and time-stratified case–crossover analyses.

2. Materials and methods

2.1. Data collection

Tianjin is a city in northeastern China, and is a directly-controlled municipality by the central government of China. Tianjin is adjacent to the Beijing city and Hebei Province, along the coast of the Bohai Gulf (39°07′ North, 117°12′ East). Tianjin has four clear seasons, with cold, windy, dry winters influenced by the vast Siberian anticyclone, and hot, humid summers due to the monsoon. It is the fifth largest Chinese city in terms of urban land area. The population in the urban area was 4.24 million in 2005.

Mortality data were obtained from the China Information System for Death Register and Report of Chinese Centre for Disease Control and Prevention (China CDC) from January 1, 2005 to December 31, 2007. The mortality data were from six urban districts of Tianjin (Heping, Hedong, Hexi, Nankai, Hebei and Hongqiao). Data on cardiovascular deaths were classified according to the International Classification of Disease, 10th revision (ICD10: I00-I99).

Daily air pollution data on particulate matter less than 10 μm in aerodynamic diameter (PM_{10}), sulfur dioxide (SO_2) and nitrogen dioxide (NO_2) were obtained from the Tianjin Environmental Monitoring Centre. Daily mean temperature and relative humidity were obtained from the China Meteorological Data Sharing Service System.

2.2. Data analysis

Applying time series analyses using generalized additive models (GAMs), we controlled for trend and season using a natural spline. In order to investigate the best possible control, the degrees of freedom per year for time were varied. We plotted the relative risks of current day's air pollution against the degrees of freedom per year for time (Peng and Dominici, 2008). The best degrees of freedom were chosen by finding the value beyond which the relative risks did not change (Peng and Dominici, 2008).

A polynomial distributed lag model was used to examine which day's exposure to air pollution had the strongest association with CVM. We examined the lagged effect for up to ten days, and used a polynomial smooth with 4 degrees of freedom (Santos et al., 2008).

As a comparison to the time series analysis we used a time-stratified case–crossover analysis. Four types of time-stratified design were used. The data were stratified by calendar month (Method A) and strata of 28 days (Method B), 21 days (Method C) and 14 days (Method D). Control days were also matched to case days by day of the week. This is why all strata used were multiples of seven days except method A.

Jackknifing is used in statistics to estimate standard errors and biases (Rothpearl, 1989). The basic principle is sub-setting the available data into many subsamples, and systematically computing estimates in the new subsamples. The bias and variance for estimates of any statistic can be calculated from these new subsamples. In this study we used jackknifing to sequentially remove strata from the data, and then the time series and case–crossover analyses were repeated. The removed strata for the time series analysis and method A were calendar month. The removed strata for method B were 28 days, for method C 21 days, and for method D 14 days. Using this method we obtained multiple estimates for the effects of air pollution on CVM. The differences between the effect estimates from the time series and case–crossover analyses were assessed using box plots and ANOVA with post hoc testing.

The conditional logistic regression used in case–crossover analysis is a special case of time series log-linear model (Lu et al., 2008; Lu and Zeger, 2007). The log-linear models were used to fit case–crossover analyses in this study. This means we can obtain the model residuals for case–crossover analyses from the log-linear models. Model residuals were examined to evaluate the adequacy of the time series and case–crossover models.

Daily mean temperature, relative humidity, public holidays, and influenza outbreaks were controlled for in all models (Braga et al., 2001). We controlled for day of the week as a categorical covariate in the time series models. Relative risks (RRs) for time series analysis, odds ratios (ORs) for time-stratified case–crossover analysis, and 95% confidence intervals (CIs) were calculated for each pollutant. All statistical tests were two-sided, and values of $P<0.05$ were considered statistically significant. R software (version 2.10.1) was used to do data analysis, the "dlnm" package was used to construct the polynomial distributed lag basis (Armstrong, 2006; Gasparrini et al., 2010), and the "mgcv" package was applied to fit the time series GAM.

3. Results

Table 1 gives descriptive statistics for the daily weather conditions, air pollutants, and mortality. The mean concentrations for PM_{10}, SO_2, and NO_2 were 105 μg/m^3, 68 μg/m^3, and 47 μg/m^3, respectively, which