

CASE-CROSSOVER DESIGN: AIR POLLUTION AND HEALTH OUTCOMES

MIECZYŚLAW SZYSZKOWICZ and NEIL TREMBLAY

Population Studies Division, Health Canada, Ottawa, ON, Canada

Abstract

Objectives: The objective of this study was to investigate variants of case-crossover design for assessing correlations between counts of health events over time and exposure to ambient air pollution. For illustrative purposes, daily emergency department (ED) visits for depression among females were considered. **Materials and Method:** Ambient nitrogen dioxide (NO₂) was used as a principal ambient air pollutant. In addition, sulphur dioxide (SO₂) and carbon monoxide (CO) were considered. Different configurations of the control periods (every 1, 2, ..., 10 days) and different forms (linear, natural splines) of meteorological factors in the applied conditional logistic regression models were used. The sequence of overlapping age intervals was defined ([0, 19], [1, 20], and so on) and each age group was analyzed separately. The results for the defined age sequences allow identifying age ranges in which the effects are strongest. **Results:** Consequently, for example, different age ranges for patients for which ED visits for depression were correlated with NO₂ and SO₂ were identified. This age-interval difference explains the very often observed phenomenon whereby two air pollutants used in one model may not show correlations with health outcomes. In this situation they affect separate age groups. The results also show dependency on number-defined control periods for the applied case-crossover technique. The opposite statistical conclusions may be generated by using different control schemas. **Conclusions:** The results support the hypothesis that ED visits for depressive disorder may be correlated with exposure to ambient air pollution.

Key words:

Air pollution, City, Depression, Emergency department, Humidity, Temperature, Nitrogen dioxide, Sulphur dioxide

INTRODUCTION

Numerous studies have linked air pollution to adverse human health effects. Notably, fine particulate matter has been linked to increased rates of hospitalization for cardiovascular and respiratory disease [1–3]. In addition, air pollution has been linked to increased rates of mortality [4,5]. The aim of this study was to consider a few methodological issues related to case-crossover (CC) design [6]. The CC methodology is an adaptation of the case-control design used in epidemiological studies. By definition of the case-crossover technique, the cases acted as their own controls on a set of predefined control time periods proximate to the time they became cases. The CC methodology

is very often used to assess the health effects of air pollution. A time-stratified approach to determine controls was widely adopted as it has been shown to produce unbiased conditional logistic regression estimates [7].

In the presented study, real data were used — mainly for realistic illustrations. The generated results are reported as odds ratios (OR) and their respective 95% confidence intervals (95% CI). These results may be used to compare with other results from different sites for the same health outcome (in this case, depression). The estimated ORs were reported for an increase in the concentration represented as an interquartile range (IQR = 75th–25th values of percentiles) of the specified air pollutant.

Received: April 27, 2011. Accepted: July 6, 2011.

Address reprint request to M. Szyszkowicz, Health Canada, 269 Laurier Avenue, Room 3-030, Ottawa, ON, K1A 0K9, Canada (e-mail: mietek.szyszkowicz@hc-sc.gc.ca).

In this study, the case-crossover method was applied to determine correlations between ambient air pollutants and emergency department (ED) visits for depression. The time unit is one day. The health data were obtained for a 10-year period from five hospitals in Edmonton, Canada. The results for ED visits for depression in Edmonton, from an analysis performed by using a different methodological approach (hierarchical clusters method), were previously published [8]. All values related to air pollutants, meteorological factors, and ED visits for depression were tabulated and presented in the mentioned publication. These values are not presented here. Again, the main purpose of this work is to highlight some issues related to the CC technique.

Recently, one specific version of the CC technique has been widely used. It is a time-stratified approach. This popular version can be called the standard method, because very often it is used in this sense. In the standard design, the controls are matched to case periods by day of the week (DOW); for the case period (day), the control periods are considered to be all the same days of the other weeks throughout the rest of the month. Consequently, in such a situation there are only 3 or 4 controls. In general, the number of controls may be increased to improve statistical stability.

MATERIALS AND METHODS

This work was inspired by the following observations. The standard CC design was applied to analyze the data related to ED visits for depressive disorder among female patients during the warm period (April–September). In the applied models, the meteorological factors, temperature and relative humidity, were adjusted in a linear form. The following results were obtained from the standard model (i.e., controls ± 7 days): for carbon monoxide (CO, IQR = 0.4 ppm), OR = 1.08, (95% CI: 0.99–1.18), p-value = 0.072; and for the nitrogen dioxide (NO₂, IQR = 12.8 ppb), OR = 1.07, (95% CI: 0.98–1.17), p-value = 0.15. In addition, for the

same data, another version of the CC method was applied in this work [9]. In this situation, control days for a particular case were selected to be every third day within the same month and year that an ED visit occurred. The DOW variable was included in the model as an indicator variable for day of the week. In this model temperature and relative humidity were included as a quadratic spline on the ‘day of visit’ and the ‘day before visit’ [4]. The obtained corresponding values for odds ratios were as follows: for CO, OR = 1.08, (95% CI: 1.00–1.18), p-value = 0.067; and for NO₂, OR = 1.10, (95% CI: 1.01–1.20), p-value = 0.04. This example illustrates that different approaches used to realize the CC technique may result in opposite conclusions on the correlations between air pollution (NO₂) and depression [10].

Statistical analysis

The unstable situation concerning the correlation between nitrogen dioxide and the numbers of ED visits for depression was intriguing. It prompted a more in-depth assessment of this relationship. For this purpose, a sequence of 67 overlapping age groups, each of length 20 years, was defined in the following way: the first age group was (0, 19), the second (1, 20), and each subsequent group was created from the current age group by shifting it by one year. In this sequence the last one was defined to be the age group (66, 85). For the elements of this sequence, the CC analysis was performed independently. Very often in an epidemiological analysis, in the domain of air health effects considered here, one specific predefined age group is chosen, usually with support of medical arguments (e.g., vulnerable patients such as the very young, the very old, or those with pre-existing medical conditions, etc.). Here, the analysis was done for 67 separate age groups, using the standard version of the case-crossover design and the CC technique with ± 3 days of controls in the same month as the corresponding case [9].

In addition, a more general approach was developed. In all the considered models the control period was restricted

to the range of one calendar month. The following model was considered:

$$\text{Pollutant} + \text{DOW} + \text{ns}(\text{Temperature, dof}) + \text{ns}(\text{Relative Humidity, dof}), \quad (1)$$

where

ns — a natural spline with specified degrees of freedom (dof).

This notation is only applied to symbolize the right-hand side of the conditional logistic regression model. The variable DOW is used to adjust for the effect of day of the week. Such a type of the CC method may be realized for different configurations of control periods. The analysis was performed using the above models with controls generated every 1, 2, ..., 9, and 10 days. Using these models the analysis was conducted for the whole period (January–December) for females and for nitrogen dioxide as a principal pollutant. The entire period has a larger sample size than that of the warm period and thus the results should be more stable. In addition, the effects of nitrogen dioxide were analyzed together with two co-pollutants: sulphur dioxide (SO₂) and carbon monoxide (CO).

RESULTS

The results are organized and presented in the form of three figures and two tables. Figure 1 represents 67 values of ORs and their corresponding 95% CIs for ED visits for depression: (top) — generated by the standard form of the case-crossover design; (bottom) — generated by the CC design with ± 3 days as the control set. The results are presented for females, for the warm months (April–September), for nitrogen dioxide. Figure 2 is analogous to Figure 1 and presents the findings for sulphur dioxide (IQR = 2.3 ppb). The main message from both figures is that the pollutants, NO₂ and SO₂, show effects for different age groups. This may impact how these pollutants work in a common model as two co-pollutants. For example, NO₂ and SO₂ may be associated with more ED visits in different time periods.

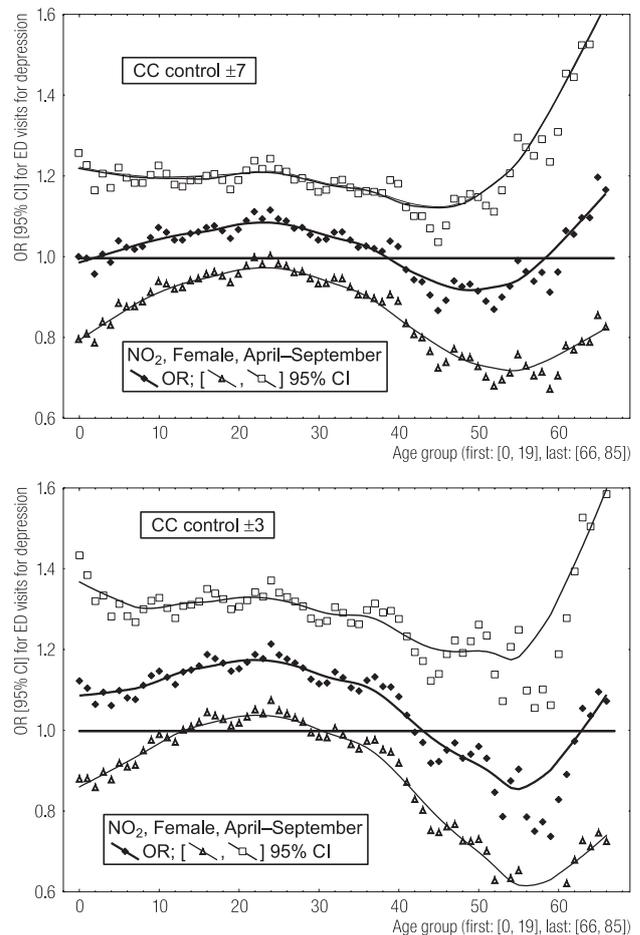


Fig. 1. ED visits for depression by age group: ORs with their corresponding 95% CIs for NO₂. Two sets of control days ± 7 and ± 3 days.

Figure 3 shows the medians and two quartiles (Q1 and Q3) for NO₂ and SO₂ concentrations for case and control periods by the overlapping age groups for female ED visits for depression. The figure represents values in IQR units for NO₂ and SO₂, respectively.

Table 1 shows the estimated ORs and their corresponding 95% CIs for ED visits for depression for different distances between control days, from 1 to 10 days. All females in the warm months (April–September) and in the whole period (January–December) were included in the analysis. Temperature and relative humidity were adjusted in the models in two forms: linear (warm season, all seasons) and natural splines (all).

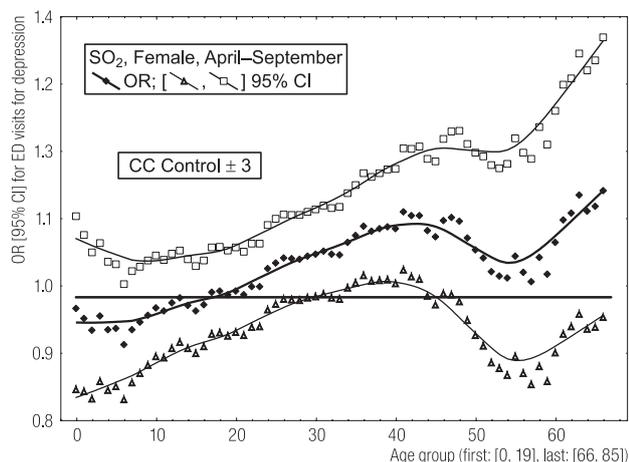
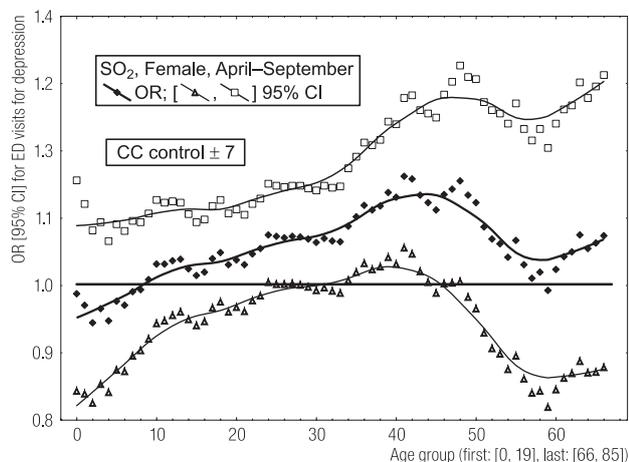


Fig. 2. ED visits for depression by age group: ORs with their corresponding 95% CIs for SO₂. Two sets of control days ±7 and ±3 days.

Table 2 reports the values of ORs and their estimated 95% CIs for ED visits for depression obtained from two-pollutant models: nitrogen dioxide with sulphur dioxide adjusted linearly and as a natural spline, and nitrogen dioxide with carbon monoxide in linear form. The controls are every 1, 2, days and so on, up to 10 days. Thus, this table is similar to Table 1 but here two co-pollutants are used in the constructed models. Nitrogen dioxide in the pair (NO₂, SO₂) does not show positive significant effects, likely because these pollutants are correlated with ED visits for different age ranges (See Figure 1 and 2). As a result, daily counts summarized ED visits for different ages

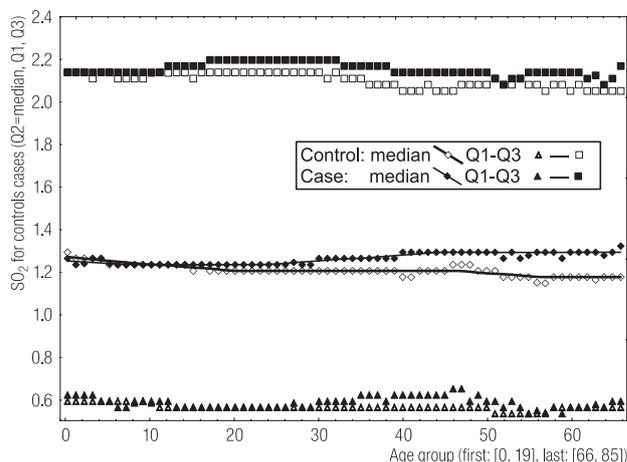
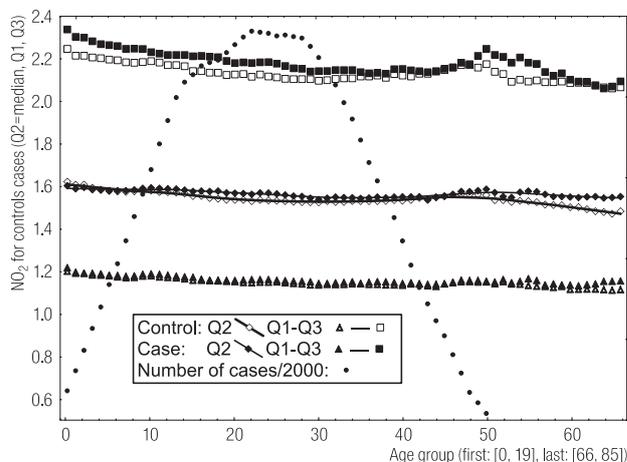


Fig. 3. Air pollutant concentrations: median and two quartiles (Q1, Q3) for cases and controls; top — nitrogen dioxide, bottom — sulphur dioxide.

were used to create a combination of the two pollutants and the corresponding cases and control periods. In this way, cases triggered by sulphur dioxide may be identified by the CC technique as controls for nitrogen dioxide, and vice versa. For example, in the warm period (See Figure 1 and 2) for NO₂ with sulphur dioxide as a co-pollutant, for the age group (25, 55), the following results were obtained: OR = 1.141 (95% CI: 1.022–1.274) with a linear form, and OR = 1.107 (95% CI: 0.991–1.237) with a spline form, The principal pollutant (NO₂) has significant correlations estimated by the models with carbon monoxide as a co-pollutant.

Table 1. Odds ratios (OR) and their corresponding 95% confidence intervals (95% CI) for an increase in interquartile range (IQR) of concentrations for NO₂ (IQR = 12.8 ppb)

±Days	Warm season		All seasons			
	linear form		linear form		spline form	
	OR	95% CI	OR	95% CI	OR	95% CI
1	1.067	0.983–1.158	1.039	0.999–1.081	1.043	1.002–1.086
2	1.068	0.983–1.159	1.041	1.000–1.083	1.044	1.003–1.087
3	1.077	0.991–1.171	1.039	0.998–1.082	1.043	1.001–1.087
4	1.082	0.994–1.178	1.045	1.002–1.089	1.049	1.006–1.094
5	1.074	0.985–1.171	1.041	0.998–1.085	1.046	1.002–1.092
6	1.063	0.974–1.161	1.038	0.996–1.083	1.041	0.997–1.086
7	1.067	0.976–1.166	1.046	1.002–1.092	1.052	1.007–1.099
8	1.107	1.011–1.212	1.050	1.005–1.098	1.055	1.009–1.103
9	1.076	0.980–1.182	1.037	0.991–1.084	1.043	0.997–1.092
10	1.035	0.940–1.139	1.030	0.983–1.079	1.034	0.986–1.084
[±6] ^a	1.081	0.995–1.175	1.058	1.015–1.102	1.062	1.019–1.107

Seasonal temperature and relative humidity are included in two forms: linear (warm, all) and spline forms (all).

Controls are used for every 1, 2, ..., 10 days (±Days) in the period of one calendar month.

^a Controls are all days in one month excluding 6 days from both sides of event day.

Table 2. Odds ratios (OR) and their corresponding 95% confidence intervals (95% CI) for ED visits for depression associated with NO₂ (IQR = 12.8 ppb) as a co-pollutant

±Days	Control					
	SO ₂ — linear (for NO ₂ +SO ₂)		SO ₂ — spline (for NO ₂ +SO ₂)		CO — linear (for NO ₂ +CO)	
	OR	95% CI	OR	95% CI	OR	95% CI
1	1.032	0.982–1.085	1.027	0.983–1.073	1.066	0.998–1.139
2	1.035	0.984–1.089	1.027	0.983–1.074	1.066	0.998–1.140
3	1.033	0.982–1.087	1.033	0.988–1.080	1.070	1.000–1.144
4	1.042	0.989–1.097	1.028	0.982–1.077	1.080	1.008–1.156
5	1.029	0.977–1.084	1.024	0.977–1.073	1.067	0.996–1.144
6	1.028	0.975–1.084	1.030	0.983–1.079	1.067	0.995–1.144
7	1.032	0.977–1.089	1.029	0.980–1.079	1.056	0.981–1.136
8	1.057	1.000–1.117	1.038	0.989–1.090	1.099	1.021–1.182
9	1.046	0.989–1.107	1.024	0.974–1.077	1.074	0.997–1.157
10	1.018	0.962–1.078	1.006	0.955–1.060	1.039	0.963–1.121
[±6] ^a	1.047	1.002–1.094	1.046	1.000–1.095	1.090	1.019–1.166

The second pollutant (SO₂, in NO₂+SO₂) is adjusted in two forms: linear and spline, and in the case of carbon monoxide (CO, in NO₂+CO), in linear form.

^a Controls are all days in one month excluding 6 days from both sides of event day.

DISCUSSION AND CONCLUSIONS

In this study, positive and statistically significant short-term effects on daily ED visits for depression for females were observed in relation to exposures to ambient air pollution. The study results show that the CC models may produce different results, as OR values and their 95% CIs (significance). It would be a good practice to execute a few runs with different forms of the CC models before disseminating the results, as the presented example for NO₂ showed that the standard version of the case-crossover design excluded this pollutant as a depression trigger for females in the warm period. The hierarchical cluster model [8] and the CC model with a larger number of the controls (every 3 days), for this case, identified NO₂ as a pollutant which exacerbates ED visits for depression.

The question remains as to which configuration of control periods to choose in a given study. Tables 1 and 2 indicate that the CC method with ± 8 days generates the largest values of ORs. Of course, this is not a universal observation and conclusion. In this study, for this configuration of control days (every 8th day), the mean value of nitrogen dioxide concentrations for a control period was the lowest among other realized configurations; thus, we had the largest contrast in levels of ambient pollutant between cases and controls. Also, this configuration (8 days) compares \pm one day to case day, with respect to day of the week. It is interesting to note that with controls defined as all days in one month with a 6-day exclusion around the case day, the case-crossover method resulted in an OR = 1.062 (95% CI: 1.019–1.107) for nitrogen dioxide. Controls were defined as all days in one month 6 days before and 6 days after the case day.

Predefined age intervals usually cut the whole age range by separated subintervals. It may happen that they split ages affected by air pollution into separately analyzed age groups, in which the ages affected (correlated to the air pollutant) will be dominated by non-affected. Consequently, the used subdivision will reduce the statistical power to detect correlations.

Different ambient air pollutants usually affect patients in different age ranges — i.e., the effects can be pollutant-specific. Accordingly, even for the same health outcome and the same age group, two pollutants in a common model may cancel out their respective effects. This situation may be recognized by constructing the sequence of overlapping age groups and performing the calculations for each age group separately. As can be seen in Figures 1 and 2, a common interval for two pollutants may be identified. In such an interval, two pollutants applied together in one model show an effect.

The limitations of this study are typical of this type of research. They include the adequacy of the model and the impact of measurement error in the exposure and outcome variables.

The results support the hypothesis that ED visits for depressive disorder may be correlated with ambient air pollution, the large majority of which arises from combustion of fossil fuels (e.g., motor vehicles).

ACKNOWLEDGEMENTS

The authors acknowledge Environment Canada for providing the air pollution data from its National Air Pollution Surveillance (NAPS) network for use in this study. The authors appreciate the efforts of Health Canada in securing these data and for funding ED data acquisition. The first author thanks Dr. Meredith Franklin for a fruitful discussion on her publication.

REFERENCES

1. Bell ML, Ebisu K, Peng RD, Walker J, Samet JM, Zeger SL, et al. *Seasonal and regional short-term effects of fine particles on hospital admissions in 202 US counties, 1999–2005*. *Am J Epidemiol* 2008;168(11):1301–10.
2. Ko FW, Tam W, Wong TW, Chan DP, Tung AH, Lai CK, et al. *Temporal relationship between air pollutants and hospital admissions for chronic obstructive pulmonary disease in Hong Kong*. *Thorax* 2007;62(9):780–5.

3. Dominici F, Peng RD, Bell ML, Pham L, McDermott A, Zeger SL, et al. *Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases*. J Am Med Assoc 2006;295(10):1127–34.
4. Samet JM, Dominici F, Curriero FC, Coursac I, Zeger SL. *Fine particulate air pollution and mortality in 20 U.S. cities, 1987–1994*. N Engl J Med 2000;343(24):1742–9.
5. Pope CA 3rd, Burnett RT, Thurston GD, Thun MJ, Calle EE, Krewski D, et al. *Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease*. Circulation 2004;109(1):71–7.
6. Maclure M. *The case-crossover design: a method for studying transient effects on the risk of acute events*. Am J Epidemiol 1991;133(2):144–53.
7. Janes H, Sheppard L, Lumley T. *Case-crossover analyses of air pollution exposure data. Referent selection strategies and their implications for bias*. Epidemiology 2005;16(6):717–26.
8. Szyszkowicz M. *Air pollution and emergency department visits for depression in Edmonton, Canada*. Int J Occup Med Environ Health 2007;21:1–5.
9. Franklin M, Zeka A, Schwartz J. *Association between PM_{2.5} and all-cause and specific-cause mortality in 27 US communities*. J Expo Scid Environ Epidemiol 2007;17:279–87.
10. Gelman A, Stern H. *The difference between “significant” and “not significant” is not itself statistically significant*. Am Stat 2006;60(4):328–31.