

## Bidirectional Case–Crossover Studies of Air Pollution: Bias from Skewed and Incomplete Waves

Jong-Tae Lee,<sup>1</sup> Ho Kim,<sup>2</sup> and Joel Schwartz<sup>3</sup>

<sup>1</sup>Department of Preventive Medicine and Public Health, College of Medicine, Yonsei University, Seoul, Korea; <sup>2</sup>Department of Epidemiology and Biostatistics and Institute of Public Health, School of Public Health, Seoul National University, Seoul, Korea;

<sup>3</sup>Environmental Epidemiology Program, Harvard School of Public Health, Boston, Massachusetts, USA

The case–crossover design compares exposures during the period of time of failure with one or more periods when failure did not occur and evaluates the potential excess risk using conditional logistic regression. In this simulation study, we applied several control sampling approaches to control for confounding by various temporal patterns of an exposure variable and evaluated the usefulness of symmetric bidirectional control strategies. We simulated true relative risks (RRs; true  $\beta = 0.001$ ) of deaths of 1.051 per 50-ppb increase of sulfur dioxide and included confounding by right- or left-skewed seasonal waves, linear long-term time trends, or a combination of both. The range of the estimated RRs from symmetric bidirectional control sampling approaches was 1.044–1.056 at either a long-term trend or any skewed seasonal wave of SO<sub>2</sub> levels, which indicated the bidirectional control sampling methods would successfully control confounding by design. The simulations with bidirectional sampling, however, show that biases may occur if waves are incomplete (20–43% underestimated RRs). In conclusion, our simulations show that the symmetric bidirectional case–crossover design can substantially control for confounding by linear long-term trends and/or seasonality of an exposure variable by design as well. However, unidirectional control sampling would fail to control confounding by those variations of air pollution. Simulation results also show that even the bidirectional case–crossover design can be biased in a situation where the exposure variable shows incomplete cyclic waves, and therefore it cannot completely control for temporal confounding. **Key words:** air pollution, case–crossover designs, control sampling strategy, epidemiologic methods. *Environ Health Perspect* 108:1107–1111 (2000) [Online 1 November 2000].

<http://ehpnet1.niehs.nih.gov/docs/2000/108p1107-1111lee/abstract.html>

Two recently published case–crossover studies on air pollution epidemiology by Neas et al. (1) and Lee and Schwartz (2) present an attractive alternative approach to analyzing the mortality effects of short-term exposure to ambient air pollution. Instead of using time-series analysis, the case–crossover study design developed by Maclure (3) was applied to evaluate the association between daily death counts and air pollution. This approach compares exposures during the period of time of death (case period) with one or more periods when death did not occur (control periods) and evaluates the potential excess risk using conditional logistic regression.

The study design requires no additional control subjects to be sampled and can control individual susceptibility by making comparisons within a subject. Therefore, this design can control for all measured and unmeasured time-invariant potential confounders. As Pope (4) described, this approach controls for potential confounding variables such as day of the week, seasonality, long-term time trends, and changes in population size and composition by design rather than by statistical modeling.

Various case–crossover designs have been proposed, involving different strategies for

selecting control information. In a case–crossover study, there are two main reasons for applying different control sampling schemes. One is to improve the relative efficiency described by Mittleman et al. (5), who showed that the empirical relative efficiency increased as the number of control periods sampled increased. The other reason is to control for exposures with temporal patterns such as long-term time trends and seasonal waves in exposure, outcome, or both (1,2,6,7). The latter problem is related to the validity of the case–crossover estimator. If the estimator were not valid, then the estimation of efficiency would not be meaningful. Navidi (6) demonstrated that the existence of systematic seasonal patterns in exposure could introduce bias into the estimates if air pollution exposure is the only source of the seasonal pattern in the outcome. He proposed a bidirectional control sampling approach, with controls randomly selected from all days before or after the event day, which eliminated the bias. In the presence of seasonal patterns in outcome, Lee and Schwartz (2) and Bateson and Schwartz (7) showed unidirectional control approaches may yield confounded risk estimates depending on the seasonal pattern or long-term time

trends of air pollution levels. Bateson and Schwartz (7) also showed that control periods needed to be close to event periods to avoid this confounding.

We begin by clearly defining “long-term time trends” or “seasonal waves.” “Long-term time trends” of exposure indicate whether the levels of exposure are tending to decrease (or increase) with calendar time. In most developed countries, the annual average levels of air pollution have decreased since air quality standards have become stricter.

The results from case–crossover studies can be quite sensitive to the selection of control periods, especially when clear time trends or cyclic waves such as seasonal waves exist. Navidi (6) illustrated how time trends in exposure can cause a bias in case–crossover studies with hypothetical examples. Lee and Schwartz (2) also showed empirically that the validity of estimates in case–crossover studies varied greatly depending on the strategy used in control sampling. Several studies suggested choosing symmetric control periods both before and after the date of death (1,2,6,7). This symmetric control sampling approach can be used under circumstances where subsequent levels of exposure are not influenced by the failure/event of interest, which is quite true in studies of air pollution and daily mortality. This new design approach was also examined thoroughly to determine whether the case–crossover design was able to control for temporal confounding patterns by design rather than through modeling (7). Symmetric bidirectional control sampling approaches with short intervals between event and control could control for temporal confounding by design (7).

Address correspondence to J-T. Lee, Department of Preventive Medicine and Public Health, College of Medicine, Yonsei University, Shinchon-Dong 134, Seodaemun-Gu, Seoul 120-752, Korea. Telephone: (822) 361-5357. Fax: (822) 392-0239. E-mail: jlee@yumc.yonsei.ac.kr

This work was financially supported by a G-7 Environmental Engineering Technology Development Project from the Department of Environment, Republic of Korea, and was also partially supported by Ministry of Health and Welfare, Republic of Korea (HMP-99-M-09-0007).

Received 23 May 2000; accepted 10 July 2000.

In the present study, we applied several control-sampling schemes to control for time trends and seasonal waves in exposure of interest (ambient air levels of sulfur dioxide). As shown in Figure 1, air pollution levels show strong seasonal waves as well as long-term time trends. To compare the methods for different control sampling approaches, we performed a simulation study. We examined two scenarios not examined by Bateson and Schwartz (7). In one scenario the seasonal pattern in SO<sub>2</sub> exposure was asymmetric; in the other scenario, data were accrued over an interval not equal to an integer number of cycles.

## Materials and Methods

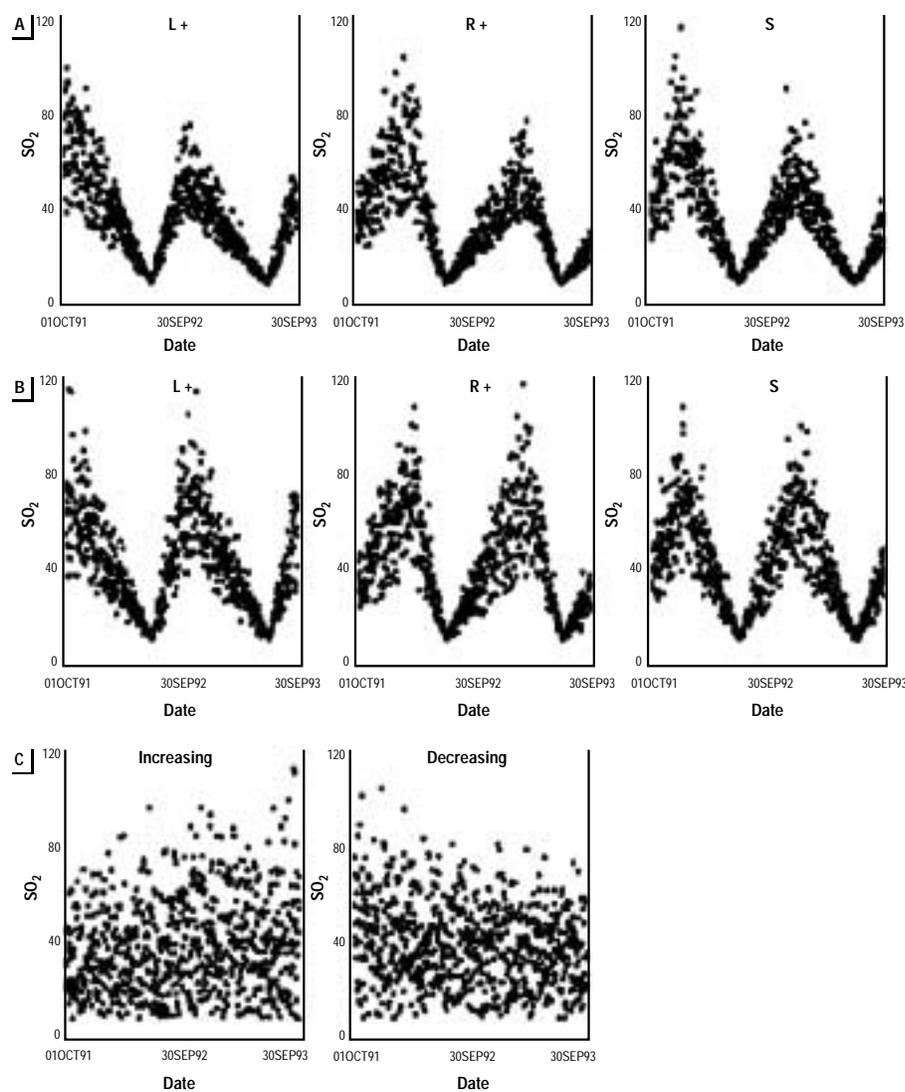
Our simulation study used death counts collected in Seoul, Korea. The National Statistics Office of Korea supplied the number of deaths occurring between 1 October 1991 and 30 September 1993, according to the day on which death occurred. In our previous study, SO<sub>2</sub> levels were significantly associated with increased all-cause mortality and showed a clear fluctuation by calendar time (2,8). The annual levels of SO<sub>2</sub> had decreased (long-term time trends). The seasonal pattern was also shown for SO<sub>2</sub> (whereas the total suspended particulates and O<sub>3</sub> seasonal waves were more scattered). It seemed clear that the periods of increasing SO<sub>2</sub> were short (steep increase in the wave), whereas the subsequent declines were gradual, over longer time periods.

Based on this pattern, we created several scenarios depending on the pattern of SO<sub>2</sub> levels over time. The fluctuation in levels could show long-term time trends, seasonal waves, or a combination of the two patterns. If there were cyclic waves (seasonal waves), the periods of increasing SO<sub>2</sub> were short (long), whereas the downward wave pattern held over longer (shorter) time periods. In some cases, it may have been true that those periods of increasing exposure levels were equal to decreasing levels. We called the shorter periods of increasing exposure levels “left heavier” or “right skewed” (Figure 1). In contrast, “right heavier” or “left skewed” was used to refer to shorter periods of decreasing exposure levels (Figure 1). Levels of exposure, on average, may have also been continuously decreasing/increasing, and we referred to this situation as long-term time trends. Of course, there could also be no long-term time trends in exposure levels. All these situations included nine scenarios, depending on the pattern of exposure levels. They can be seen in Figure 1, except for the case of no trends at all.

To compare the methods described earlier in this paper, we performed a simulation study. First, we generated sets of SO<sub>2</sub> data from 1 October 1991 to 30 September 1993. Data were artificially created so that they had

the nine patterns described earlier. The details of a simulation procedure are given in Appendix 1. The error terms with variance proportional to the SO<sub>2</sub> values were added to the SO<sub>2</sub> values so that the data looked naturally distributed. Figure 1 shows the generated daily SO<sub>2</sub> values for nine patterns. After the SO<sub>2</sub> value was simulated, the death count was simulated from a Poisson distribution using SAS RANPOI function (SAS Institute, Cary, NC) with a mean parameter  $\lambda = 82 \times \exp(0.001 \times \text{SO}_2)$ , where 82 was the approximate mean death count in Seoul. If we assume there were 10,000,000 people at risk in Seoul, the baseline hazard was  $82/10,000,000 = 8.2 \times 10^{-6}$  for all people in Seoul, and the true value of  $\beta$  was assumed

to be 0.001. This corresponds to the true relative risk of approximately 1.051 per 50 ppb, which is in close agreement with the relative risk of mortality due to SO<sub>2</sub> as estimated by Lee and Schwartz (2). For each data set of death counts, the estimated value,  $\hat{\beta}_i$ , was calculated by the conditional logistic regression method using SAS PROC PHREG (9), and this process was repeated 1,000 times. Then the following were calculated: a) the mean of estimated values ( $\hat{\beta}_i, i = 1, 2, \dots, 1,000$ ), b) the sample standard deviation calculated from the estimated values ( $\hat{\beta}_i, i = 1, 2, \dots, 1,000$ ), and c) the root mean-squared error (RMSE) of  $\hat{\beta}$ , which is the square root of the sample mean of squared errors  $[(\hat{\beta}_i - \beta)^2, i = 1, 2, \dots, 1,000]$ .



**Figure 1.** Simulated daily levels of SO<sub>2</sub> (ppb) from 1 October 1991 to 30 September 1993. (A) SO<sub>2</sub> levels were created as if they had both long-term time trends (decreasing with calendar time overall) and seasonal waves. (B) A seasonal wave only (no long-term time trends of SO<sub>2</sub> levels). (C) A long-term time trend only (no seasonal waves). Left heavier (L+, right skewed) indicates that the periods of increasing levels are short, whereas the downward wave pattern holds over longer time periods. Right heavier (R+, left skewed) indicates the reverse relation to L+. Symmetric (S) indicates that the periods of increasing SO<sub>2</sub> levels or downward wave patterns are similar.

The mean-squared error (MSE) of  $\hat{\beta}$  is defined by Equation 1.

$$\begin{aligned} \text{MSE}(\hat{\beta}) &= E[(\hat{\beta} - \beta)^2] \\ &= [E(\hat{\beta}) - \beta]^2 + E[\hat{\beta} - E(\hat{\beta})]^2 \\ &= \text{Bias}^2 + \text{Var}(\hat{\beta}), \end{aligned} \quad [1]$$

where  $\text{Bias} = E(\hat{\beta}) - \beta$  and  $\text{Var}(\hat{\beta}) = E[\hat{\beta} - E(\hat{\beta})]^2$ .

As we can see in Equation 1, the MSE is the sum of the squared bias and the squared standard error (variance) of  $\hat{\beta}$ . For illustrational purpose, we use RMSE rather than MSE because RMSE has the same dimension of bias and standard error. In Tables 1–3, each row represents the sampling design results retrospective with one control, retrospective with two controls, bidirectional with two controls, bidirectional with four controls, prospective with one control, and prospective with two controls, respectively. In our study, controls were selected 1 week

and 2 weeks before and/or after the death date for the corresponding sampling designs in order to be limited to a similar part of the same season with the case.

## Results

We can gain some insight into the nature of the bias for unidirectional sampling approaches by examining the time trends in simulated data. Under one scenario, air pollution levels were created to have time trends only, without any cyclic wave such as seasonal pattern. The results can be seen in Table 1. The first three columns of Table 1 present, for each estimator, summaries from the 1,000 replications across the simulated data sets in which the exposure levels decrease with calendar time. The next three columns present the same information in which the exposure levels increase with calendar time. By the nature of simulated data sets, on average, the days when death occurred had lower air pollution levels than their corresponding

unidirectional retrospective control days in the first situation and vice versa in the second situation. The results of unidirectional control are biased in the presence of increasing or decreasing time trends in exposure, similar to the results of Navidi (6) and Bateson and Schwartz (7). The bias is greater with 2 control days. In this example, we can conclude that bidirectional sampling approaches are nearly unbiased and their RMSEs, combined measure of bias and efficiency, are smaller, while others are considerably biased and RMSEs are larger if time trends exist.

We also simulated the data sets to have seasonal waves. Interpretation of this analysis is more complicated than the analysis of data sets for time trends only. However, it should be mentioned that we obtained the least biased estimates with smallest RMSE from bidirectional sampling approaches. There were two scenarios where data sets included seasonal waves. One included both seasonal waves and long-term time trends in exposure.

**Table 1.** Comparison of case–crossover estimators by various sampling approaches in a situation where only long-term time trends of SO<sub>2</sub> levels exist.

Estimator (true $\beta = 0.001$ ) <sup>a</sup>	Decreasing by time (year)			Increasing by time (year)			Constant (year)		
	Mean	SE ( $\times 10^{-3}$ )	RMSE ( $\times 10^{-3}$ )	Mean	SE ( $\times 10^{-3}$ )	RMSE ( $\times 10^{-3}$ )	Mean	SE ( $\times 10^{-3}$ )	RMSE ( $\times 10^{-3}$ )
$\beta_{U1-}$	0.00075	0.4503	0.5153	0.00137	0.4406	0.5736	0.00105	0.4311	0.4336
$\beta_{U2-}$	0.00059	0.3835	0.5594	0.00167	0.4138	0.7855	0.00111	0.4125	0.4271
$\beta_{B2}$	0.00109	0.3345	0.3456	0.00097	0.3357	0.3367	0.00094	0.3635	0.3689
$\beta_{B4}$	0.00107	0.2700	0.2788	0.00098	0.3633	0.3637	0.00107	0.3540	0.3613
$\beta_{U1+}$	0.00141	0.3907	0.5649	0.00057	0.4919	0.6523	0.00097	0.4315	0.4323
$\beta_{U2+}$	0.00156	0.3091	0.6433	0.00028	0.5162	0.8881	0.00115	0.4523	0.4752

RMSE, root mean-squared error.

<sup>a</sup> $\beta_{U1-}$ , unidirectional retrospective with one control;  $\beta_{U2-}$ , unidirectional retrospective with two controls;  $\beta_{B2}$ , bidirectional with two controls;  $\beta_{B4}$ , bidirectional with four controls;  $\beta_{U1+}$ , unidirectional prospective with one control;  $\beta_{U2+}$ , unidirectional prospective with two controls.

**Table 2.** Comparison of case–crossover estimators by various sampling approaches in a situation where only seasonal waves of SO<sub>2</sub> levels exist.

Estimator (true $\beta = 0.001$ ) <sup>a</sup>	Right heavier			Left heavier			Symmetric		
	Mean	SE ( $\times 10^{-3}$ )	RMSE ( $\times 10^{-3}$ )	Mean	SE ( $\times 10^{-3}$ )	RMSE ( $\times 10^{-3}$ )	Mean	SE ( $\times 10^{-3}$ )	RMSE ( $\times 10^{-3}$ )
$\beta_{U1-}$	0.00102	0.5899	0.5902	0.00139	1.1340	1.2061	0.00083	0.7086	0.7292
$\beta_{U2-}$	0.00084	0.5179	0.5408	0.00120	1.1191	1.1375	0.00049	0.6236	0.8063
$\beta_{B2}$	0.00102	0.4744	0.4749	0.00090	0.5325	0.5417	0.00098	0.5142	0.5145
$\beta_{B4}$	0.00095	0.4055	0.4091	0.00094	0.4201	0.4247	0.00099	0.4109	0.4110
$\beta_{U1+}$	0.00096	0.5933	0.5949	0.00060	1.1144	1.2119	0.00123	0.6918	0.7302
$\beta_{U2+}$	0.00113	0.5543	0.5693	0.00091	0.9013	0.9057	0.00161	0.6950	0.9254

RMSE, root mean-squared error.

<sup>a</sup> $\beta_{U1-}$ , unidirectional retrospective with one control;  $\beta_{U2-}$ , unidirectional retrospective with two controls;  $\beta_{B2}$ , bidirectional with two controls;  $\beta_{B4}$ , bidirectional with four controls;  $\beta_{U1+}$ , unidirectional prospective with one control;  $\beta_{U2+}$ , unidirectional prospective with two controls.

**Table 3.** Comparison of case–crossover estimators by various sampling approaches in a situation where both long-term time trends (decreasing with calendar time overall) and seasonal waves of SO<sub>2</sub> levels exist.

Estimator (true $\beta = 0.001$ ) <sup>a</sup>	Right heavier			Left heavier			Symmetric		
	Mean	SE ( $\times 10^{-3}$ )	RMSE ( $\times 10^{-3}$ )	Mean	SE ( $\times 10^{-3}$ )	RMSE ( $\times 10^{-3}$ )	Mean	SE ( $\times 10^{-3}$ )	RMSE ( $\times 10^{-3}$ )
$\beta_{U1-}$	-0.00012	0.8576	1.4074	-0.00173	0.7447	2.8308	-0.00009	0.7729	1.3369
$\beta_{U2-}$	-0.00057	0.8161	1.7723	-0.00324	0.7961	4.3131	-0.00083	0.6882	1.9554
$\beta_{B2}$	0.00090	0.5784	0.5978	0.00100	0.6535	0.6535	0.00099	0.5969	0.5869
$\beta_{B4}$	0.00090	0.4806	0.4912	0.00097	0.5810	0.5817	0.00087	0.5340	0.5504
$\beta_{U1+}$	0.00200	0.8562	1.3187	0.00374	0.7090	2.8291	0.00214	0.7428	1.3573
$\beta_{U2+}$	0.00255	0.8166	1.7554	0.00530	0.7282	4.3599	0.00293	0.6340	2.0279

RMSE, root mean-squared error.

<sup>a</sup> $\beta_{U1-}$ , unidirectional retrospective with one control;  $\beta_{U2-}$ , unidirectional retrospective with two controls;  $\beta_{B2}$ , bidirectional with two controls;  $\beta_{B4}$ , bidirectional with four controls;  $\beta_{U1+}$ , unidirectional prospective with one control;  $\beta_{U2+}$ , unidirectional prospective with two controls.

The other included only seasonal waves in exposure, without linear time trends. The former is more realistic in air pollution studies. Air pollution levels in most developed or developing countries have exhibited long-term time trends, as well as seasonal patterns. In most countries, the ambient levels of air pollutants have been steadily decreasing since ambient air quality standards were strengthened. It is also a well-known phenomenon that certain kinds of air pollution show cyclic fluctuation by seasons. Tables 2 and 3 show that, for the case of symmetric seasonal patterns in exposure, the bidirectional method gives an essentially unbiased estimate with smallest RMSE. This replicates the results of Bateson and Schwartz (7). However, for the case of left- or right-heavy patterns in exposure, the bidirectional approach shows more bias and more RMSE, although in other settings it gives less biased estimates than unidirectional sampling. In comparing RMSE, it is interesting to note that Table 1 shows similar values of RMSEs for both bidirectional methods. However, Tables 2 and 3 show that the bidirectional method with four controls gives smaller RMSEs than the bidirectional method with two controls. This suggests that in a real situation, the bidirectional method with

many controls will give the less biased and more efficient estimator if there is seasonal variation or long-term trend.

Our simulation analysis also showed that symmetric control periods may not work well if the exposure has cyclic waves. We observed that the estimates with symmetric control periods could be biased if the data sets with a cyclic wave were incomplete (Table 4). Here, a complete cyclic wave means that a data set whose number of concave vertex points (peaks) and convex vertex points (valleys) are the same. A complete wave does not necessarily mean the multiples of a complete year's data. As shown in Figure 2, data from summer to summer or from winter to winter have different numbers of concave vertex points and convex vertex points, so they are incomplete waves. For the incomplete cyclic wave, the symmetric control sampling approaches could also be biased. For entire exposure data sets, the estimator could be biased if the number of convex vertex points was greater (or less) than the number of concave vertex points (Table 4). The RRs estimated with bidirectional case–crossover design were underestimated by 20–43% compared with the true RR (Table 4). Figure 2 depicts the situation described above. This

analysis suggests that we need to look at the entire exposure pattern thoroughly if the exposure has a cyclic wave such as seasonal patterns. If the wave is not complete, then the estimates with symmetric bidirectional sampling approaches could be biased as well.

## Discussion

In environmental epidemiologic studies, the interpretation of the results has been criticized because the data may be subject to an unmeasured confounding bias. The size of association between air pollution and daily mortality is also small enough that the difference may be considered due to bias. To overcome all those difficulties in conducting environmental research, researchers need to focus on the use of improved environmental monitoring techniques, the amount and precision of health and exposure data, the application of advanced statistical techniques, and the study design.

Formal time-series analysis using Poisson regression has been the primary statistical approach to such studies. Recently, advanced time-series modeling techniques have been developed to provide better control for potential confounders. For example, generalized additive models (GAM) that use nonparametric smoothing have allowed for highly flexible fitting of seasonal wave and long-term time trends, as well as nonlinear associations with weather variables such as air temperature and humidity. Most previous mortality studies have dealt with death counts rather than individual deaths.

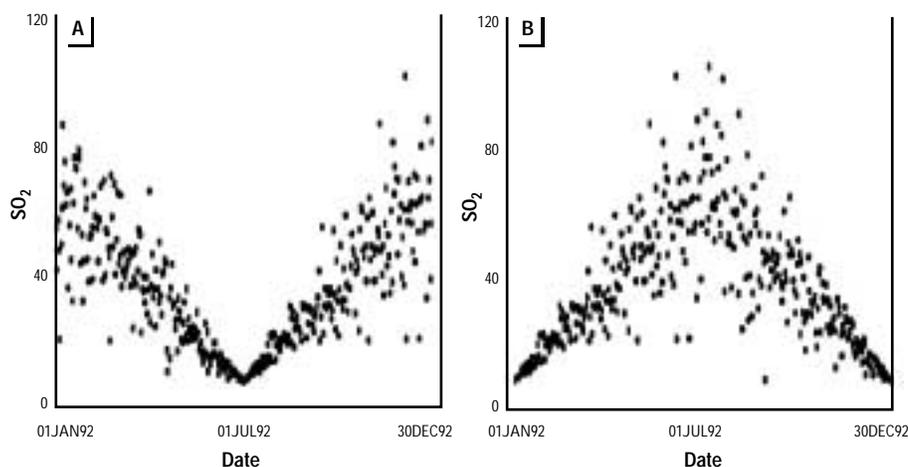
As Pope (4) pointed out, case–crossover studies have some drawbacks. Due largely to the loss of information from control periods that cannot be included in the analysis, the case–crossover approach has been believed to have lower statistical efficiency. If a bidirectional sampling approach with a few control days (which are selected within the limited periods, for example within 2 weeks) had been applied for 1 year of data, then there would not be more than 7.7% (4 weeks divided by 1 year) information loss. This indicated that deaths could not be analyzed for a 4-week period (2 weeks from the beginning and 2 weeks from the end) each year. The relative efficiency will not be substantially lowered if the study periods are extended with a few control days. However, Bateson and Schwartz (7) found 66% efficiency in their simulations when controls were randomly selected. The major drawback of a case–crossover design was that the results could be sensitive to the selection of control periods, especially when clear time trends exist. Previous studies have shown that bidirectional approaches are needed when there are seasonal patterns in exposure (6) or in both exposure and outcome (7).

**Table 4.** Comparison of case–crossover estimators by various sampling approaches in a situation where seasonal waves of SO<sub>2</sub> levels are not complete.

Estimator (true $\beta=0.001$ )	Cup-shaped pattern			Cap-shaped pattern		
	Mean	SE ( $\times 10^{-3}$ )	RMSE ( $\times 10^{-3}$ )	Mean	SE ( $\times 10^{-3}$ )	RMSE ( $\times 10^{-3}$ )
$\beta_{U1-}$	0.00265	0.8163	1.8421	0.00093	0.7672	0.7696
$\beta_{U2-}$	0.00239	0.7233	1.5693	0.00098	0.6654	0.6653
$\beta_{B2}$	0.00066	0.6986	0.7783	0.00078	0.5943	0.6338
$\beta_{B4}$	0.00081	0.6128	0.6426	0.00057	0.5263	0.6766
$\beta_{U1+}$	-0.00069	0.7945	1.8702	0.00109	0.7538	0.7586
$\beta_{U2+}$	-0.00014	0.7194	1.3458	0.00127	0.6553	0.7075

RMSE, root mean-squared error.

<sup>a</sup> $\beta_{U1-}$ , unidirectional retrospective with one control;  $\beta_{U2-}$ , unidirectional retrospective with two controls;  $\beta_{B2}$ , bidirectional with two controls;  $\beta_{B4}$ , bidirectional with four controls;  $\beta_{U1+}$ , unidirectional prospective with one control;  $\beta_{U2+}$ , unidirectional prospective with two controls.



**Figure 2.** Simulated daily levels of SO<sub>2</sub> (ppb) in which the seasonal pattern is not complete. (A) A cup-shaped pattern where a concave vertex point exists. (B) A cap-shaped pattern where a convex vertex point exists.

Although our results confirm these earlier findings, we also found that for left-heavy or right-heavy exposure patterns, even the approaches recommended in those studies can result in biased estimates. Further, if there is an incomplete number of cycles in the exposure data, bias may occur. Given that the actual pattern of air pollution variation used to be sinusoidal, and some pollutants such as ozone are often monitored for only part of the year, substantial caution is called for in applying the case–crossover approach to time–series data. At this stage, we cannot suggest any quantitative procedure to determine whether the data are appropriate to case–crossover design. This could be a further research topic to clarify. However, a couple of methods can be mentioned: one is to apply various control sampling strategies and compare the estimates from those sampling approaches, another is to empirically observe the cyclic waves of raw data. The efficiency of case–crossover design estimators has been shown to be lower than that of ordinary Poisson regression estimators (7). A future study could focus on the issue of the efficiency problem of case–crossover design.

It cannot be determined whether the case–crossover design is better or worse than various advanced time–series approaches in air pollution epidemiology. However, it is clear that this alternative approach can be put to use and that it adds supporting evidence in understanding the association between daily mortality and the current lower levels of air pollution. A clearer understanding of the relationship between the method of selecting controls and case–crossover estimators is needed.

#### REFERENCES AND NOTES

1. Neas LM, Schwartz J, Dockery D. A case–crossover analysis of air pollution and mortality in Philadelphia. *Environ Health Perspect* 107:629–631 (1999).
2. Lee JT, Schwartz J. Reanalysis of the effects of air pollution on daily mortality in Seoul, Korea: a case–crossover design. *Environ Health Perspect* 107:633–636 (1999).
3. Maclure M. The case–crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol* 133:144–153 (1992).
4. Pope CA III. Mortality and air pollution: associations persist with continued advances in research methodology. *Environ Health Perspect* 107:613–614 (1999).
5. Mittleman MA, Maclure M, Robins JM. Control sampling strategies for case–crossover studies: an assessment of relative efficiency. *Am J Epidemiol* 142:91–98 (1995).
6. Navidi W. Bidirectional case–crossover designs for exposures with time trends. *Biometrics* 54:596–605 (1998).
7. Bateson TF, Schwartz J. Control for seasonal variation and time trend in case–crossover studies of acute effects of environmental exposures. *Epidemiology* 10:539–544 (1999).
8. Lee JT, Shin D, Chung Y. Air pollution and daily mortality in Seoul and Ulsan, Korea. *Environ Health Perspect* 107:149–154 (1999).
9. SAS Institute. *SAS/STAT Software: Change and Enhancements through Release 6.12*. Cary, NC: SAS Institute Inc., 1997.

## Appendix

A sine curve should be used to simulate air pollution data to capture the cyclic seasonal pattern. Because the real shapes of SO<sub>2</sub> plots are sharper at the edges, we decided to use several lines to simulate SO<sub>2</sub> values rather than one sine curve. Another reason for using combinations of lines is that it is easier to create left- or right-skewed data. We controlled the ranges of the date, slopes, and intercepts of the lines so that we had the shapes we wanted. For long-term trend only (no seasonal waves) simulation, only one mean line was used to generate SO<sub>2</sub> values. The formula for the simulation was

$$SO_2 = a \times \text{day} + b + N(0,1) \times (c \times \text{day} + d),$$

where day is the number of days after 1 October 1991 ( $1 \leq \text{day} \leq 731$ ) and  $N(0,1)$  are the random numbers generated from standard normal distribution using NORMAL function of SAS version 6.12 (9). Values of  $a$ ,  $b$ ,  $c$ ,  $d$  are given in Table A-1. For the simulations for seasonal waves, five mean lines were used to generate SO<sub>2</sub> values. The whole range of the day ( $1 \leq \text{day} \leq 731$ ) was divided into five pieces, depending on the patterns of the skewness, and a mean line was used for each piece of the range. The cut points for five pieces of day ranges are given in Table A-1. And in the individual piece of the range, a similar formula was used

$$[SO_2 = a \times \text{day} + b + N(0,1) \times (c \times \text{day} + d)].$$

To give the linear trend, 0.7 was multiplied at the second cycle (see Table A-1). For all simulations, if the simulated value is negative, then the value was deleted and a new random number was generated again until a positive value was generated.

**Table A-1.** Coefficients to generate SO<sub>2</sub> values.

Shape	Range of the day	Coefficient			
		<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
A long-term time trend only (no seasonal waves)					
Increasing trend	1–731	0.0153214	32	0.0109439	22.8571
Decreasing trend	1–731	–0.0153214	32	–0.0109439	22.8571
Cyclic (seasonal) wave with no time trend					
Symmetric cyclic wave (S)					
	1–92	0.3347	36.555	0.1046	36.55
	93–274	–0.3347	97.73	–0.1046	97.73
	275–458	0.3347	–85.50	0.1046	–85.80
	459–639	–0.3347	200.09	–0.1046	220.08
	640–731	0.3347	–246.25	0.1046	–208.15
Right heavier (or left-skewed) wave (R+)					
	1–164	0.2389	27.87	0.0747	6.85
	165–274	–0.5578	158.87	–0.1743	47.78
	275–530	0.2389	–59.53	0.0747	20.47
	531–640	–0.5578	362.81	–0.1743	111.52
	641–731	0.2389	–146.94	0.0747	–47.78
Left heavier (or right-skewed) wave (L+)					
	1–18	0.5578	57.01	0.1743	15.95
	19–274	–0.2389	71.46	–0.0747	20.47
	275–384	0.5578	–146.94	0.1743	–47.78
	385–640	–0.2389	158.87	–0.0747	47.78
	641–731	0.5578	–350.89	0.1743	–111.52
A decreasing long-term time trend with a seasonal wave					
	Same as S, R+, and L+, respectively.		0.7 × coefficients of S, R+, and L+, respectively		

$$SO_2 = a \times \text{day} + b + N(0,1) \times (c \times \text{day} + d)$$