

A Case-Crossover Analysis of Air Pollution and Mortality in Philadelphia

Lucas M. Neas,^{1,2,3} Joel Schwartz,^{1,3} and Douglas Dockery^{1,3}

¹Environmental Epidemiology Program, Department of Environmental Health; and ²Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA; ³Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

This study reassessed Schwartz and Dockery's analysis of daily mortality from nonexternal causes among residents of Philadelphia, Pennsylvania, over 8 years, from 1973 to 1980 [*American Review of Respiratory Disease* 145:600–604 (1992)]. A Poisson regression analysis using the same model found that a 100- $\mu\text{g}/\text{m}^3$ increment in the 48-hr mean concentration of total suspended particulates (TSP) was associated with increased all-cause mortality [rate ratio = 1.069; 95% confidence interval (CI), 1.043–1.096] after adjustment for quadratic trend, season, year, previous day's mean temperature, dew point, winter temperature, and indicators of hot (temperature > 80°F) and humid days (dew point > 66°F). Critics suggested that time-varying factors such as season and day of week were not sufficiently controlled in this analysis and subsequent studies in other locations. We used a conditional logistic regression analysis with a case-crossover design to reanalyze the data, with air pollution in the prior and subsequent weeks to the day of death serving as referent periods. The case-crossover approach controls for season and day of week by design rather than modeling. We found that a 100- $\mu\text{g}/\text{m}^3$ increment in the 48-hr mean level of TSP was associated with increased all-cause mortality [odds ratio (OR) = 1.056; CI, 1.027–1.086] after adjustment for the same weather variables as above. Similar associations were observed for deaths in individuals over 65 years of age (OR = 1.074; CI, 1.037–1.111) and for deaths due to cardiovascular disease (OR = 1.063; CI, 1.021–1.107). The current case-crossover analysis confirms the general conclusion of the previous Poisson regression analysis of an association of TSP with daily mortality in Philadelphia, Pennsylvania. *Key words:* air pollutants, case-crossover, mortality. *Environ Health Perspect* 107:629–631 (1999). [Online 24 June 1999] <http://ehpnet1.niehs.nih.gov/docs/1999/107p629-631neas/abstract.html>

The most common technique for modeling the covariation of environmental exposures and counts of health events over time has been Poisson regression analysis (1,2). An approach that examines associations over time rather than over space is attractive because many potential covariates are controlled by design, rather than by regression modeling. Because the same population is examined repeatedly under varying exposure conditions, time invariant characteristics such as sex and race are not potential confounders. Covariates that have secular trends or vary seasonally may be weakly correlated with environmental exposures, but the potential for these covariates to confound the short-term health effects of particulate matter may be controlled by modeling time trends and season. These advantages have allowed Poisson regression analyses of daily time series to identify airborne particles and ozone as predictors of deaths and hospital admissions in many studies around the world (3). Recently, even more sophisticated analytical techniques have been introduced for trend and seasonal adjustment, culminating in the introduction of generalized additive models (2). However, some critics have argued that the particulate matter associations are not robust to alternative techniques for the control of trend and season (4,5).

Another issue in these Poisson regression analyses has been the lack of a full

enumeration and characterization of the population at risk. To omit the offset term from the Poisson regression model, the investigator must assume that the population at risk is very large relative to the daily number of events and that the composition and size of the population at risk does not covary with the exposure of interest. The later assumption may not be fully met whenever the susceptible portion of the total population at risk may be increased by the cumulative effects of prior exposures or decreased by the adverse effects of prior exposures (harvesting).

The case-crossover design avoids both problems by restructuring the analysis (6). The case-crossover design is an adaptation of the case-control design in which cases serve as their own controls. A subject's characteristics and exposures at the time of a health event (case period) are compared with another time period when that subject was a noncase (control period). Each risk set consists of a single individual who crosses over to different exposure levels in the intervals between the control and case periods. Because each subject serves as his or her own control, this approach controls for stable subject-specific covariates by design. These matched sets of a case period and one or more control periods may be analyzed using conditional logistic regression. The size of the population at risk is not an issue with this design.

The case-crossover design also controls many time-varying confounders by design because the case and control periods in each risk set are separated by a relatively small interval of time. This time interval may be only a few hours in some cardiovascular research (7,8) or several days in air pollution research. Factors that vary slowly over a longer time scale, such as trend, season, and smoking status, are essentially the same in both periods and therefore do not confound the health effects of more rapidly varying factors such as air pollution. Weather factors such as temperature and precipitation may covary on the same time scale as air pollution and so must be controlled in the analysis. Long-term variations in the prevalence of chronic health problems cannot confound, but chronic disease may be an indicator of a sensitive subpopulation.

We have used the case-crossover technique to reassess Schwartz and Dockery's analysis of the association of total suspended particulates (TSP) with daily mortality from nonexternal causes among residents of Philadelphia, Pennsylvania, over 8 years, from 1973 to 1980 (9).

Materials and Methods

Site description. In 1980, Philadelphia had 1,688,710 residents. The regional transport of combustion-related particles and oxidants was the primary source of ambient air pollutants, but major local sources included particles from mobile sources and a complex of petroleum refineries south of the city.

Pollutant and meteorological measurements. Pollution data were retrieved from the national aerometric data banks of the U.S. Environmental Protection Agency (10). Daily monitoring for TSP was conducted at two population-based monitors (1501 East Lycoming Street and 500 South Broad Street) supplemented by sampling every sixth day at several additional sites. Citywide

Address correspondence to J. Schwartz, Department of Environmental Health, Harvard School of Public Health, 665 Huntington Avenue, Boston, MA 02115 USA. Telephone: (617) 432-1245. Fax: (617) 277-2382. E-mail: joel@sparc6b.harvard.edu

L.N. was supported in part by U.S. Environmental Protection Agency cooperative agreement CR-821762. J.S. and D.D. are supported in part by NIEHS grant ES 0002, and J.S. is supported in part by NIEHS grant ES 07410.

Received 11 January 1999; accepted 23 March 1999.

means for TSP were calculated for each day from the available monitoring data. As in the previous paper (9), the TSP data were averaged over a 48-hr period (midnight to midnight) ending on the specified date. Meteorological data included the 24-hr mean temperature on the previous day and the 24-hr mean dew-point temperature on the day of death.

Daily mortality. All deaths of Philadelphia residents were extracted from detailed mortality tapes prepared by the National Center for Health Statistics (NCHS) for the calendar years 1973–1980. Deaths from accidental causes [International Classification of Diseases, 9th Revision (ICD-9), codes ≥ 800] were excluded, as were all deaths that occurred outside of the city. Daily mortality was summarized as the total, by age (< 65 years and ≥ 65 years) and by principal underlying cause of death [cardiovascular disease (ICD-9 390–448) and cancer (ICD-9 140–208)]. One or more deaths from each of these causes occurred on 74% of the days during the study period.

Statistical methods. The case-crossover design uses conditional logistic regression models to analyze the exposure odds for the case period as compared with the control period. The standard rate model for conditional logistic regression is

$$E(x_{i,j} / y_{i,j} | Z_i) = \frac{1}{\theta_{Y,j}} \theta_{0,j} \rho^{Z_i} \\ = \frac{\rho^{Z_i}}{\sum \rho^{Z_i}}$$

where $x_{i,j}$ is the number of cases in the j th matched risk set in the i th covariate pattern, $y_{i,j}$ is the number of controls, Z_i is the independent variables for the i th covariate pattern, $\theta_{Y,j}$ is the sampling rate for the controls, $\theta_{0,j}$ is the baseline hazard, and ρ^{Z_i} is the rate ratio for the i th covariate pattern. The sampling rate and the baseline hazard rate in a conditional logistic model are not directly estimated because the model conditions on these parameters.

The control periods are generally selected to control relevant time-varying confounders. For ambient air pollutants, the case period is the 48-hr period ending at midnight on the day of death. The reference or control periods are selected to control confounding by day of week, that is 7, 14, or 21 days before the case period and 7, 14, or 21 days after the case period. Symmetric control periods around the case period were chosen to control for the long-term time trend in the data. As with any matched design, multiple reference periods may be used in the analysis. All of the deaths occurring on a given day share

the same exposures in both the case period and control period. A detailed discussion of control sampling strategies for case-crossover studies has been previously published (11).

We analyzed the daily mortality counts by a case-crossover model using the SAS conditional logistic regression procedure (12). The case periods and control periods paired by date were modeled as a function of the covariates, with weights equal to the number of deaths on that day. The covariates included the 48-hr pollutant concentration ending on the day of the death, the 24-hr mean temperature on the previous day, the 24-hr mean dew point temperature, an indicator for hot days with average temperatures greater than 80°F, an indicator of humid days with dew points greater than 66°F, and an interaction between winter season and current 24-hr mean temperature.

Results

Meteorology and air pollutant concentrations. Complete data on TSP, temperature, dew point, and relative humidity were available for 2,669 days over the 8-year period of 1973–1980. The 24-hr average temperatures ranged from 4 to 88°F, with an average of 55°F. The 24-hr average temperatures exceeded 80°F on 153 days (5.7%) and these days were characterized as hot days. During the winter months (December through February), the 24-hr temperatures averaged 33°F. The 24-hr mean dew point temperatures ranged from -17 to 76°F, with an average of 43°F. The 24-hr average dew point temperatures exceeded 66°F on 346 days and these days were characterized as humid days. Dew point temperatures were highly correlated with temperatures, with a Pearson's correlation coefficient (r) of 0.88.

The 48-hr average TSP levels ranged from 24.5 to 233 $\mu\text{g}/\text{m}^3$, with an average of 76.9 $\mu\text{g}/\text{m}^3$ and an interquartile range of 32 $\mu\text{g}/\text{m}^3$. A gradual decline in TSP air pollution was observed over the study period, but the decline was modest (-2.8 $\mu\text{g}/\text{m}^3/\text{year}$).

Description of daily mortality. Over the 2,669 days with complete air monitoring

data, there were 127,614 deaths from nonexternal causes among Philadelphia residents: 83,172 among residents <65 years of age (65%), 58,178 due to cardiovascular disease (46%), and 28,023 due to cancer (2%). Total daily mortality ranged from 26 to 86 deaths, with an interquartile range of 11 deaths. Mortality declined by 0.7 deaths/year and showed an annual cycle [coefficient of determination (r^2) = 0.14] in a model with two pairs of sine and cosine terms with periods of 365 and 120 days (9).

Poisson regression models. A reanalysis of the data using Poisson regression models replicated the results reported in the earlier paper (9). In the Poisson models, trend was modeled with linear and quadratic functions and year and season were modeled with indicator variables. A 100- $\mu\text{g}/\text{m}^3$ increment in TSP was associated with total mortality in models controlling for trend, year, season, and the meteorological covariates [rate ratio = 1.069, 95% confidence interval (CI), 1.043–1.096]. Meteorological variables were modeled as in the original paper for comparability.

Case-crossover regression models. Using six reference periods (7, 14, and 21 days before and after the case period), a 100- $\mu\text{g}/\text{m}^3$ increment in TSP was associated with total mortality [odds ratio (OR) = 1.056; CI, 1.027–1.086] after adjustment for the 24-hr mean temperature on the previous day, the 24-hr mean dew point temperature, an indicator for days with average temperatures > 80°F, an indicator of humid days, and an interaction between winter season and current 24-hr mean temperature. A model with four symmetric reference periods 7 and 14 days around the case period produced a similar result, but a model with only two symmetric reference periods of 7 days around the case period produced a stronger association (Table 1).

To test the adequacy of the case-crossover design to control for seasonal variation, we included two pairs of sine and cosine terms with periods of 365 and 120 days. With symmetric reference periods, the

Table 1. Adjusted^a odds ratios for the association of daily mortality with a 100 $\mu\text{g}/\text{m}^3$ increment in the 48-hr average of total suspended particulates (TSP), Philadelphia, Pennsylvania, 1973–1980.

Control periods	Without seasonal adjustment ^b	With seasonal adjustment ^b
Before and after case period		
7 days	1.105	1.104
7 and 14 days	1.050	1.048
7, 14, and 21 days	1.056	1.055
Before case period only		
7 days	1.123	1.106
7 and 14 days	1.073	1.047
7, 14, and 21 days	1.072	1.041

The common case period for exposure was a 48-hr average TSP ending on the day of death.

^aAdjusted for yesterday's 24-hour average temperature, today's dew point, winter temperature, and an indicator of hot days. ^bAdjustment for season with four sine and cosine functions with periods of 365 and 120 days.

association of TSP with mortality was not altered by adjustment for season (Table 1). When only prior reference periods were used in the analysis, a larger pollution effect was seen before, but not after, this cyclical adjustment (Table 1).

The crude association of TSP and mortality without adjustment for weather (OR = 1.086) was substantially larger than the adjusted association (OR = 1.056). The association was not further reduced by the inclusion of an indicator of humid days. The relationship between TSP and mortality did not depart greatly from linearity when TSP was modeled adding a quadratic term ($\chi^2_1 = 2.8$) or by a natural cubic spline with knot points at the upper and lower quartiles ($\chi^2_4 = 5.9$).

A larger effect was seen for deaths in persons ≥ 65 years of age (OR = 1.074; CI, 1.037–1.111) and for deaths due to pneumonia (OR = 1.076; CI, 0.918–1.260) and to cardiovascular disease (OR = 1.063, 95% CI 1.021–1.107) in the model with six reference periods. Cancer mortality was not associated with TSP (OR = 1.004; CI, 0.945–1.066).

Discussion

In earlier Poisson regression analyses (9), the association between TSP and daily deaths in Philadelphia was robust to the method of modeling time trends, seasonal patterns, and weather. Similar results were found using generalized additive models, which use non-parametric smoothing to fit nonlinear functions of time and weather (3). Schwartz (3), Schwartz and Dockery (9), and Kelsall et al. (13) found that these results were insensitive to model specification. However, Li and Roth (4) and Moolgavkar et al. (5) have challenged these reports. Further, the degree of sensitivity may depend on the correlations between particulate air pollution, season, and weather, which vary from city to city. Hence the results in other locations may be less robust to model specification. A methodology that controls for such patterns by design is clearly a valuable contribution.

The case-crossover approach has the potential to control for some of these factors by design by comparing the current exposure period in the 48 hr immediately before death with referent exposure periods in the same season and close in time. However, most case-crossover studies use a prior reference period. In that case, the case event always follows the control period, and there is perfect confounding with time trends, if present. Because time trends are usually present in mortality series, symmetrical reference periods are needed (14). Symmetrical reference periods close together in time seem likely to control for season as well as trend. By using symmetric 7-, 14-, or 21-day reference

periods, the case-crossover model eliminates by design the confounding effects of long-term trend and day of week. In this analysis, the symmetric reference periods appear to control for seasonal variation as well. Further, because the case and control periods are limited to a similar part of the same season, the range of variation of the weather variables is more limited than in the Poisson regressions and therefore may be easier to control by modeling weather variables. We believe that this demonstrates the utility of the case-crossover approach in reducing the dependence of conclusions about air pollution associations on modeling assumptions.

It is important to control for season and day-of-week effects (or at least weekend/weekday contrasts) for several reasons. There is considerable seasonality in mortality that remains even after control for weather. This is due to unmeasured factors such as time spent outdoors, viability of viruses, etc., which vary seasonally. Because air pollution also has seasonal variability, this has the potential to confound the pollution associations if not controlled. In addition, season probably modifies the relationship between ambient concentrations of particles and personal exposure to particles of ambient origin. Time spent outdoors and window opening are two obvious reasons. In addition to this 365 day cycle, behavior and exposure change substantially between weekdays and weekends. A case-crossover analysis with control periods chosen as multiples of 7 days is an elegant way to accomplish these goals.

We have confirmed the previous results of Schwartz and Dockery (9), Schwartz (3), Kelsall et al. (13), and Samet et al. (15) that the association of particulate air pollution with increases in daily mortality in Philadelphia are not sensitive to change in modeling techniques. Unlike Poisson regression analyses, the case-crossover model controls for trend and season by design and limits potential confounding by weather due to the reduced range of variation in weather factors between the case and control periods. A disadvantage to the case-crossover approach is that the model for total mortality has a larger standard error in the case-crossover analysis (0.0142) than in the Poisson analysis (0.0126). This loss of statistical power will increase whenever there are some days that have no deaths. The lower statistical power of the case-crossover design has been previously noted (16).

The analysis here has been restricted to comparing the results using one pollutant. There are multiple pollutants in the air in Philadelphia, and those associations have been well documented (13,15). They are not repeated here because our goal is to

demonstrate the methodology of case-crossover designs, not repeat analyses that have been done before.

An advantage of the case-crossover approach is that it may be implemented with any statistical software that supports conditional logistic regression, whereas generalized additive models are only available in more advanced and difficult packages, such as S-Plus [Statistical Sciences Corporation, Seattle, WA (17)]. More importantly, the case-crossover approach in principle allows the investigator to utilize additional data on each individual death. Additional death certificate information such as exact age, contributing cause of death, and location of death, might be useful as covariates and effect modifiers in the case-crossover setting.

REFERENCES AND NOTES

- Schwartz J, Spix C, Touloumi G, Bacharova L, Barumamzadeh T, le Tertre A, Piekarksi T, Ponce de Leon A, Ponka A, Rossi G, et al. Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. *J Epidemiol Community Health* 50(suppl 1):S3–S11 (1996).
- Schwartz J. Air pollution and daily mortality in Birmingham, Alabama. *Am J Epidemiol* 137:1136–1147 (1993).
- Schwartz J. Air pollution and daily mortality: a review and meta-analysis. *Environ Res* 64:36–52 (1994).
- Li YZ, Roth HD. Daily mortality analysis by using different regression models in Philadelphia County, 1973–1990. *Inhal Toxicol* 7:45–58 (1995).
- Moolgavkar S, Luebeck EG, Hall TA, Anderson EL. Air pollution and daily mortality in Philadelphia. *Epidemiology* 6:476–484 (1995).
- Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol* 133:144–153 (1991).
- Mittleman MA, Maclure M, Toffler GH, Sherwood JB, Goldberg RJ, Muller JE. Triggering of acute myocardial infarction by heavy exertion: protection against triggering by regular exertion. *N Engl J Med* 329:1677–1683 (1993).
- Mittleman MA, Maclure M, Sherwood JB, Mulry R, Toffler GH, Jacobs SC, Friedman R, Benson H, Muller JE. Triggering of acute myocardial infarction onset by episodes of anger. *Circulation* 92:1720–1725 (1995).
- Schwartz J, Dockery DW. Increased mortality in Philadelphia associated with daily air pollution concentrations. *Am Rev Respir Dis* 145:600–604 (1992).
- Nehls GJ, Akland GG. Procedures for handling aerometric data. *J Air Pollut Control Assoc* 23:180–184 (1973).
- 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).
- SAS. SAS/STAT User's Guide, Version 6, Vol 2. 4th ed. Cary, NC: SAS Institute Inc., 1989.
- Kelsall JE, Samet JM, Zeger SL, Xu J. Air pollution and mortality in Philadelphia, 1974–1988. *Am J Epidemiol* 146:750–762 (1997).
- Bateson T, Schwartz J. Control for seasonal variation and time trend in case-crossover studies of acute effects of environmental exposures. *Epidemiology* (in press).
- Samet JM, Zeger SL, Kelsall J, Xu J, Kalkstein L. Does weather confound or modify the association of particulate air pollution with mortality? *Environ Res* 77:9–19 (1998).
- 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).
- Venables WN, Ripley BD. *Modern Applied Statistics with S-Plus*. New York: Springer-Verlag Inc., 1994.