ORIGINAL ARTICLE

The effects of particulate air pollution on daily deaths: a multi-city case crossover analysis

J Schwartz

....

....

Correspondence to:
Associate Professor
J Schwartz, Harvard
School of Public Health,
Exposure, Epidemiology,
and Risk Program, 401
Park Drive, Suite 415
West, Boston, MA 02215,
USA; jschwrtz@hsph.
harvard.edu

Background: Numerous studies have reported that day-to-day changes in particulate air pollution are associated with day-to-day changes in deaths. Recently, several reports have indicated that the software used to control for season and weather in some of these studies had deficiencies.
Aims: To investigate the use of the case-crossover design as an alternative.
Methods: This approach compares the exposure of each case to their exposure on a nearby day, when they did not die. Hence it controls for seasonal patterns and for all slowly varying covariates (age, smoking, etc) by matching rather than complex modelling. A key feature is that temperature can also be controlled by matching. This approach was applied to a study of 14 US cities. Weather and day of the week were controlled for in the regression.
Results: A 10 $\mu g/m^3$ increase in PM_{10} was associated with a 0.36% increase in daily deaths from internal causes (95% CI 0.22% to 0.50%). Results were little changed if, instead of symmetrical sampling of control days the time stratified method was applied, when control days were matched on temperature, or when more lags of winter time temperatures were used. Similar results were found using a Poisson regression, but the case-crossover method has the advantage of simplicity in modelling, and of combining matched strata across multiple locations in a single stage analysis.
Conclusions: Despite the considerable differences in analytical design, the previously reported associations of particles with mortality persisted in this study. The association appeared quite linear. Case-crossover designs represent an attractive method to control for season and weather by matching.

The case-crossover design, introduced by Maclure[1] in 1991, represents an attractive method to investigate the acute effects of an exposure. It has been used, for example, to investigate triggers of myocardial infarction.[2] In recent years, it has been applied to the analysis of the acute effects of environmental exposures, especially air pollution.[3–6] In the case-crossover approach, a case-control study is conducted whereby each person who had an event is matched with their self on a nearby time period where he or she did not have the event. The subject's characteristics and exposures at the time of the case event are compared with those of a control period in which the event did not occur. Each risk set consists of one individual as that individual crosses over between different exposure levels in the interval between the two time periods. These matched pairs may be analysed using conditional logistic regression. Multiple control periods may be used.

Applied to the association of air pollution with risk of death, the approach has several advantages. First, it clarifies a key feature of the study of acute response to air pollution. Because in this analysis each subject serves as their own control, the use of a nearby day as the control period means that all covariates that change slowly over time, such as smoking history, age, body mass index, usual diet, diabetes, etc, are controlled for by matching.

The second advantage involves the method of control for seasonal variations in mortality risk. The other possible technique to analyse the association of short term changes in air quality with short term changes in the risk of death or hospital admission has been to collapse the data to daily counts, and use Poisson regression of the daily data.[7] Because these regressions make comparisons across the full range of data, including multiple years, it is necessary to control for season and long term time trends. While several approaches have been taken to control for the seasonal patterns in the

data,[8 9] in 1993 generalised additive models[10] were applied to these analyses,[11] and that technique quickly became the standard in subsequent studies.[12 13] These models are attractive because they use smooth curves to control for season (and weather).

Non-parametric smoothing is attractive when one believes a non-linear association exists with a covariate, since it is more flexible than parametric approaches.[10] However, recent studies have reported problems with the algorithms that implement these models.[14] The standard errors of the parametric terms, including the hypothesis variables, are not estimated correctly.[15] Because too small estimates of within location standard errors leads to larger estimates of heterogeneity (the between location variance), recent reanalyses of multi-location studies did not show bias in the estimated standard errors of the combined effect estimates.[16 17]

These problems have encouraged reanalysing previous studies to confirm whether the reported associations still hold. Natural splines are a possible alternative, and have been used previously,[11] but they have some sensitivity to the locations of the knots for the splines. Details of spline models have been published elsewhere,[10 11] but in essence, they divide a continuous variable into a set of discrete ranges, and fit separate polynomials in each range. The boundary points of the ranges are called knots. Natural splines have recently been applied to a reanalysis of the National Mortality and Morbidity Air Pollution Study (NMMAPS), a large multi-city study of particulate air pollution and daily deaths.[14] There has been a continuing debate over how many degrees of freedom are appropriate to use to control for season, and switching from non-parametric smoothing to natural splines does not resolve this issue. More recently, similar debates have emerged about how many degrees of freedom to use for weather.

Main messages

- Case-crossover designs allow assessment of the short term effects of air pollution without the use of complex models to control for season and weather.
- They also facilitate the combination of data across study locations and examination of effect modification.
- Using this approach in a study of 14 cities, a similar association of particles with daily deaths was found as in more traditional studies.
- The particle associations are robust to method.

Policy implications

- The associations between particles and mortality risk are unlikely to be confounded by weather and season, and are robust to analytical method.
- Substantial public health benefits are likely to result from controlling the sources of this pollution.

The large number of studies of daily changes in air pollution and deaths has been a basis for tighter air pollution standards in both the United States and Europe. The recent questions over the reliability of these estimates have, therefore, considerable public health importance.

The case-crossover design controls for seasonal variation, time trends, and slowly time varying confounders by design because the case and control periods in each risk set are separated by a relatively small interval of time. Bateson and Schwartz[18][19] showed that by choosing control days close to event days, even very strong confounding of exposure by seasonal patterns could be controlled by matching in the case-control approach. Also, because the analysis is of matched strata of days for each individual, it is straightforward to combine events from multiple locations in a single analysis. The difference in seasonal patterns from city to city has prevented this approach in multi-city studies using Poisson regression. This makes the approach an attractive alternative to the Poisson models. While Bateson and Schwartz[19] have shown that the power is lower in the case-crossover approach, this is less of a concern in a large multi-city study. Because the case-crossover approach focuses on individual events, rather than daily counts, it also makes examination of effect modification more straightforward.

While it is straightforward to sample control days in a manner that removes seasonal confounding, there can be a subtle selection bias in these analyses.[19–21] For example, days before the first event serve as control days, but cannot serve as event days, and occasional days with missing data for exposure during the event series can further increase selection bias. Further, Lumley and Levy[20] have pointed out that the selection of controls is not independent sampling, since the case day always falls in the middle. This introduces a small selection bias as well. Lumley and Levy[20] suggest a time stratified approach, where the control days for a given subject are randomly selected from all days in the same month of the same year. Bateson and Schwartz[19] suggested that maintaining symmetrical control sampling would provide better seasonal control, and demonstrated an adjustment method that gives unbiased estimates. A recent simulation study showed that both approaches give unbiased estimates and unbiased coverage probabilities.[22] I have applied both approaches to a multi-city study of particulate air pollution and daily deaths in a study of 14 US cities.

For comparison, I have also analysed those cities using Poisson regression of daily counts, with penalised splines[23] used to control for season and weather. Penalised splines do not suffer the standard error problem of the earlier generalised additive model approaches.

DATA AND METHODS

Most cities in the USA only monitored PM_{10} once every six days. This can create difficulties in finding control days close to event days. To avoid this, I studied 14 US cities with daily monitoring schedules: Birmingham, AL; Boulder, CO; Canton, OH; Cincinnati, OH; Columbus, OH; Chicago, IL; Colorado Springs, CO; Detroit, MI; Minneapolis/St Paul, MN; New Haven, CT; Pittsburgh, PA; Provo-Orem, UT; Seattle, WA; and Spokane, WA. I chose the metropolitan county containing each city except for Minneapolis and St Paul, where both metropolitan counties were combined and analysed as one city.

Daily mortality

Deaths in the metropolitan county containing each city were extracted from tapes prepared by the National Center for Health Statistics (NCHS) for the calendar years 1986 (when PM_{10} monitoring began to be phased in by the US EPA) to 1993. Deaths from accidental causes (International Classification of Diseases 9th Revision, ICD9 $\geqslant$800) were excluded, as were all deaths that occurred outside of the city. While deaths were available from 1986, in some cases PM_{10} monitoring did not begin until later. This results in some variation in the study period by city.

Exposure and weather data

Daily measurements of mean temperature and relative humidity were obtained from the nearest National Weather Service Surface Station for each county (EarthInfo CD NCDC Surface Airways, EarthInfo Inc., Boulder, CO).

Air pollution data for PM_{10} were obtained from the US Environmental Protection Agency's Aerometric Retrieval System (AIRS). Many of the cities have more than one monitoring location, requiring a method to average over multiple locations. This study uses an algorithm previously reported.[24] To ensure that the exposure measure represented general population exposure and not local conditions affecting only the immediate vicinity of a given monitor, the correlations among all monitors in each county were computed. Monitors within the lowest 10th centile of the correlation across all counties were excluded. While all locations had daily monitoring for PM_{10} by at least one monitor, some monitors only measure PM_{10} one day in three or one day in six, and different monitors have different means and standard deviations. I did not want the daily pollution value to change from day to day because of changes in which monitors reported, as opposed to differences in actual ambient levels. In each city the daily mean among monitors for each pollutant was calculated using an algorithm that accounted for differences in the annual mean and the standardised deviations of each monitor as follows. The daily standardised deviations for each monitor on each day were averaged; these were then multiplied by the standard deviation of all of the monitor readings for the entire year, and added back in the annual average of all of the monitors.[13] I used the air pollution concentration the day before each death as the exposure variable, because the NMMAPS[13][14] study has found that to be the most predictive single day exposure, and presents that exposure metric in most of its publications. The NMMAPS study could not

analyse multi-day exposures because most of their cities only measured PM_{10} one day in six. The use of the same exposure metric facilitates comparisons between the studies.

Analytical strategy

Several analyses were conducted. The basic analysis used conditional logistic regression to analyse the data in each city, in a case-crossover design, using the Bateson and Schwartz method to choose control days.[19] I chose this because our simulations showed it did slightly better in the face of strong seasonal patterns and time trends.

Matched strata were constructed for each subject (that is, death), consisting of the event day (day of death) and 18 matched control days. These days were chosen to be the days 7–15 days before the event day, and the days 7–15 days after the event day. Control days were not chosen closer to the event day to avoid serial correlation in the pollution and mortality data. Control days were chosen symmetrically about the event day because Bateson and Schwartz[18] showed that symmetric control days were needed to control for long term time trends (if present). Navidi[25] pointed out that bi-directional sampling is needed to avoid some biases in the case-control study, and does not present any conceptual difficulties as long as the inactivity of the subject after death does not affect the air pollution concentrations. The adjustment method of Bateson and Schwartz[19] was used to address

the small selection bias pointed out by Lumley and Levy.[20] Briefly, this consists of creating a pseudo-data set with the same exposure variables as in the real data, but with one death on every day. By construction, there can be no association between pollution and the risk of death in this dataset. The estimated coefficient obtained from this pseudo-data set was used as an estimate of the bias, and subtracted from the naïve estimate obtained analysing the real data.

In all analyses, I controlled for day of the week, temperature, and relative humidity. Temperature may be non-linearly related to the risk of death, and so I used regression splines to control for temperature on the day of death and the day before death. These splines used 3 degrees of freedom each. Relative humidity was controlled for in a similar manner.

The first analysis used a two stage approach. A city specific regression was fit using the matched strata from each city. The log odds ratios from those 14 analyses were then combined using the iterative maximum likelihood algorithm of Berkey and coworkers,[26] allowing for heterogeneity in effect across city. In this analysis, the splines for temperature could have different coefficients in different studies. The random effects estimate was used in this case, whether or not the random variance component was significant, because the significance test for that component is weak, and I wanted to be sure to incorporate heterogeneity.

Table 1 Descriptive statistics showing dates of study in each city, and the 25th, 50th, and 75th centiles of the environmental variables in each city

City	Study dates (available PM_{10})	Temperature (°C)	Relative humidity	PM_{10} (µg/m³)	Non-accidental deaths
Birmingham	1/4/87–31/12/93	11	62	20	60996
		18	71	31	
		24	80	46	
Boulder	1/1/86–31/12/93	3	40	17	7655
		11	52	25	
		19	65	35	
Canton	1/1/89–24/12/93	2	66	19	13677
		11	74	26	
		19	82	34	
Chicago	1/3/88–24/12/93	2	62	23	267147
		11	70	33	
		19	79	46	
Cincinnati	2/8/88–31/12/93	4	64	23	41496
		13	72	32	
		21	81	44	
Colorado Springs	1/7/87–24/12/93	2	39	18	12663
		11	51	23	
		18	66	31	
Columbus	1/1/86–31/12/93	3	64	20	28907
		12	70	28	
		21	80	39	
Detroit	1/5/86–24/12/93	2	64	21	145195
		11	71	32	
		19	79	49	
Minneapolis/St Paul	1/4/87–24/12/93	−1	60	17	75529
		9	69	24	
		19	78	35	
New Haven	1/5/87–31/12/91	3	57	17	32401
		12	67	26	
		20	77	38	
Pittsburgh	1/1/87–24/12/93	3	61	19	178649
		12	70	30	
		20	79	47	
Provo-Orem	1/4/87–31/12/93	3	39	21	7697
		12	52	30	
		21	72	45	
Seattle	1/1/86–24/12/93	7	67	18	83538
		11	77	27	
		16	85	39	
Spokane	1/1/86–24/12/93	2	49	23	12964
		8	68	36	
		16	84	57	

Table 2 Correlation between PM_{10} and other environmental variables in the study locations

City	Temperature (°C)	Relative humidity
Birmingham	0.26	−0.3
Boulder	−0.07	−0.23
Canton	0.42	−0.16
Chicago	0.36	−0.3
Cincinnati	0.44	−0.00
Colorado Springs	−0.34	−0.11
Columbus	0.41	−0.07
Detroit	0.37	−0.14
Minneapolis/St Paul	0.29	−0.35
New Haven	0.05	−0.15
Pittsburgh	0.45	−0.23
Provo-Orem	−0.41	0.33
Seattle	−0.22	−0.11
Spokane	−0.01	−0.19

Because the control days are chosen close to the event day in the case-crossover analysis, the range of variation of temperature, and the range of its effects, is lower than in other study designs. This suggested that I could aggregate all of the strata together, and analyse the association with air pollution across all 14 cities in a single model. This is an attractive feature of the study design, and constituted the second analysis. It is equivalent to assuming no heterogeneity in response across cities, however.

In the third analysis, I used the time stratified method of Lumley and Levy[20] to select control days for each person who died. In addition to matching on season, I also matched on temperature. That is, I chose control days to be days in the same month and year as the death occurred, that also had the same temperature (rounded to degree Celsius). In these analyses, I continued to control for humidity on the same and the previous day, and for temperature on the previous day, using regression splines. Both two stage and single stage analyses were done using the time stratified control selection method as well.

To facilitate comparison with more traditional Poisson regression analysis, I reduced the data to counts of daily deaths, and performed Poisson regressions in each city. These models used penalised splines with 4 degrees of freedom per year for season and 3 degrees of freedom for each weather

Table 3 Percent change in risk of death for a 10 $\mu g/m^3$ increase in PM_{10} by city, and overall, using either a two stage or single stage analysis

City	% change	95% CI
Birmingham	−0.01	−0.58 to 0.56
Boulder	2.59	0.66 to 4.56
Canton	−0.02	−1.67 to 1.65
Chicago	0.39	0.14 to 0.64
Cincinnati	0.32	−0.52 to 1.15
Colorado Springs	0.35	−1.01 to 1.74
Columbus	0.61	−0.59 to 1.81
Detroit	0.47	0.15 to 0.78
Minneapolis/St Paul	0.63	0.06 to 1.21
New Haven	0.27	−0.63 to 1.19
Pittsburgh	0.30	−0.05 to 0.65
Provo-Orem	0.26	−0.77 to 2.59
Seattle	0.14	−0.32 to 0.61
Spokane	0.32	−0.20 to 1.21
Overall two stage	0.36	0.22 to 0.50
Overall single stage	0.33	0.19 to 0.46

Analyses are matched on subject and season, and control for weather (using splines) and day of the week.

Table 4 Percent change in the risk of death for a 10 $\mu g/m^3$ increase in PM_{10} in 14 US cities

Method	% increase	95% CI
More winter temperature lags*		
Two stage	0.39	0.23 to 0.56
One stage	0.32	0.19 to 0.46
Time stratified with temperature matching†		
Two stage	0.39	0.19 to 0.58
One stage	0.53	0.34 to 0.72
Poisson regression‡	0.40	0.18, 0.62

*More lags of winter temperature (lags 2–4).
†Time stratified approach to matching on season and temperature (rounded to the nearest degree Celsius.
‡Poisson regression analysis with penalised splines to control for season and weather.

variable (today's and yesterday's temperature and humidity), as well as day of the week dummy variables. The use of a pre-specified fixed number of degrees of freedom in a penalised spline model reduces it to a ridge regression, with well developed methods for estimating standard errors, unlike the problems with generalised additive models.

To test the shape of the dose-response relation, I replaced the linear term for PM_{10} with the indicator variables for days when concentrations were between 15 and 25 $\mu g/m^3$, between 25 and 34 $\mu g/m^3$, between 35 and 44 $\mu g/m^3$, and 45 $\mu g/m^3$ and above. The days with concentrations below 15 $\mu g/m^3$ served as the reference level. This model was fit using the single stage method, combining strata across all cities

RESULTS

Table 1 shows the 25th, 50th, and 75th centiles of the distribution of air pollution and weather in each of the 14 locations, as well as the total number of deaths studied in each location. Weather was only modestly correlated with PM_{10} in these locations, as shown in table 2.

In the two stage analysis, I found a significant association between PM_{10} and the odds of dying. The magnitude of the association was a 0.36% increase in the risk of death per 10 $\mu g/m^3$ increment of PM_{10} (95% CI 0.22% to 0.50%). Combining all the strata and analysing in one stage had little impact on the estimate (0.33% increase, 95% CI 0.19% to 0.46%). Table 3 shows the individual city results. There was no evidence for heterogeneity in the association ($\chi^2 = 12.17$ on 13 df, p = 0.51), but a random variance component was kept to be conservative.

Table 4 shows the results of further sensitivity analyses. Matching on same day temperature (and continuing to control for humidity and previous day's temperature) had little impact on the effect estimates, which were in fact slightly larger than those found without matching. Further, the results were not noticeably different from those found using a more traditional Poisson regression analysis.

Figure 1 shows the results of the analysis using categories of PM_{10} exposure to look at the dose-response curve. There was little evidence of deviation from linearity.

DISCUSSION

Substantively, this analysis confirms that using parametric regression techniques which avoid the software problems in generalised additive model software, one still finds a significant association between daily levels of PM_{10} and daily deaths. It extends that finding by: (a) doing so using a very different approach, indicating a robustness of the results to type of modelling; and (b) in particular using an approach
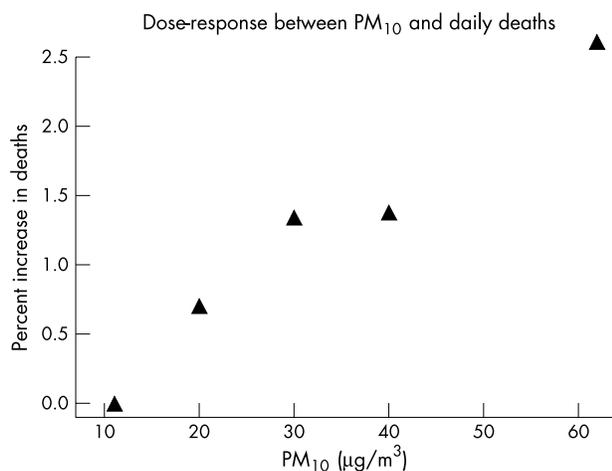
Dose-response between PM$_{10}$ and daily deaths

Figure 1 shows the percent increase in the risk death on days with PM$_{10}$ concentrations in the ranges of 15–24 μg/m^3, 25–34 μg/m^3, 35–44 μg/m^3, and 45 μg/m^3 and greater, compared to a reference of days when concentrations were below 15 μg/m^3. Risk is plotted against the mean PM$_{10}$ concentration within each category.

that avoids having to model seasonal patterns at all, obviating arguments about the complexity of those models.

Methodologically, it shows an approach that has certain advantages over the previous methods applied to multi-city studies. While case-crossover analyses have been reported in single city studies,[3–6 19] this is the first application to multi-city studies in air pollution. The results of the primary analysis are indistinguishable from the results of a Poisson regression applied to the same data. Wherein lie the potential advantages?

First, case-crossover methods allow the possibility of avoiding arguments about how many degrees of freedom and what functions are necessary to control for season, since seasonal control is done by matching. It also makes the method of seasonal control more accessible to readers less familiar with the literature on splines and smoothing. Matching within two weeks is more intuitive than choosing a number of degrees of freedom per year for seasonal control. Splines and other flexible functions can fit the data in ways not expected. For example, consider a study interested in the immediate effects of ambient temperature on mortality risk. Figure 2 shows the results of fitting a 7 degree of freedom per
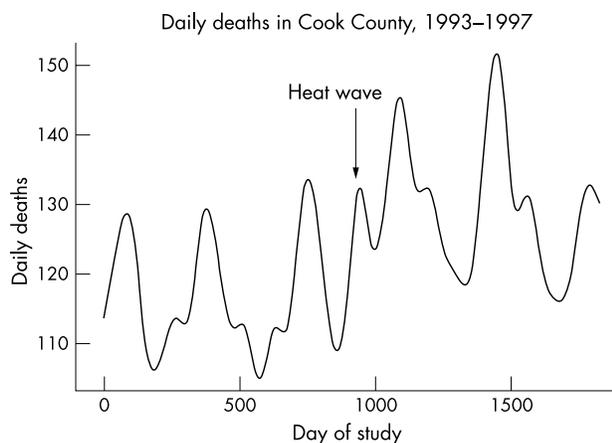
Daily deaths in Cook County, 1993–1997

Figure 2 The seasonal pattern of daily deaths in Chicago in the years 1993–97 predicted by a 7 degree of freedom spline. With this many degrees of freedom the seasonal term is starting to explain part of the 1995 Chicago heat wave, which lasted less than a week.

year spline model to explain seasonality for daily deaths in Chicago during the years 1993–97. In the summer of 1995 there was a serious heat wave in Chicago, which was responsible for an estimated 400 early deaths over a six day period. One can see that the 7 df/year seasonal model is fitting a peak of mortality in the middle of the usual summer trough in 1995. That is, 7 degrees of freedom per year is sufficient to start picking up not merely season, but also the effects of a six day weather episode. This is in contrast to many people's intuition that it only removes patterns on the order of two months or longer from the data. A standard method of attributing deaths to heat waves is to count the excess over the seasonally expected during the heat wave period. Clearly, this peak in the summer months attributable to "season" will reduce the number of excess deaths that can be attributed to the coincident heat wave by that method. A case-crossover analysis of the effects of temperature that makes sure to choose control days outside of the heat wave period would easily avoid this problem in identifying the temperature effects.

Second, as shown here, for air pollution studies, it is possible to match on temperature as well as season. This offers great advantages. It obviates the arguments about how to control for same day temperature. Also when matching on season and temperature, one implicitly controls for all interactions between those variables.

Temperature is not the only weather variable, and may exert effects with some lags. However, Braga and colleagues[27] and Cuireiro and colleagues[28] have shown that the largest weather effect is on the same day. Matching can control that portion of the weather's effects. Further, while there is no explicit matching on lagged temperature, days matched to within two weeks of time and with the same temperature do not differ greatly in temperatures on the previous days. For example, across these 14 cities, the mean standard deviation of the previous day's temperature within strata (case and control days), after matching by current day's temperature, was 2.2°C. Hence lagged temperature is indirectly controlled to a substantial extent by matching. The lack of change in the risk estimate for particulate air pollution in these cities comparing matching on temperature with regression modelling to control for temperature indicates that the models adequately controlled for temperature before matching.

An additional advantage of the case-crossover design is the ability to combine information across cities in a single stage model. Because strata of event and control days within subject are matched on season (and weather), it is possible to combine these strata in a single analysis. One example of the possibilities this engenders is the analysis of the shape of the dose-response relation presented in fig 1. While this paper presented a simple approach, more complex analyses, include spline models and smoothing, are available, and in this application need no longer be combined across cities in a two stage model. Multilevel modelling of effect modification by city characteristics is easily incorporated in the single stage model. Effect modification by individual characteristics may also be examined. In the past, this has been done by stratified Poisson regression, such as the identification of diabetes as a potential modifier of the effect of PM$_{10}$ on hospital admissions for heart disease,[29] but identification of characteristics can be more easily incorporated into the case-crossover approach.

Nevertheless, the case-crossover approach does have important limitations. One key limitation is the difficulty in examining associations over longer time periods than a few days. This limits the methods ability to examine either harvesting or associations with longer term exposure. In addition, while Lumley and Levy[20] have shown a proper conditional logistic likelihood with time stratified sampling,

and Schwartz and coworkers[22] have shown both unbiased estimates and unbiased coverage probabilities (and hence confidence intervals), more remains to be learned about the theoretical properties of this approach.

The examination of dose-response in this analysis shows little evidence for a deviation from non-linearity. However, it is important to understand that the range in exposure was not enormous in this study. For example, annual average PM_{10} concentrations in Athens exceed the highest category in fig 1. Hence non-linearities at higher concentrations cannot be excluded by this analysis. It is also interesting to note the pattern of correlations between PM_{10} and weather variables in table 2. The Midwestern cities all had moderate, positive correlations between temperature and PM_{10}. This is likely due to the importance of sulphate particles in that region. Sulphate levels, and levels of other long range transported particles, tend to increase on hot days because of increased photochemistry. In contrast, in the northwest part of the country, the correlation was negative, and smaller. This region is less impacted by sulphate. Hence differences in these correlations may reflect different sources and atmospheric processes contributing to exposure.

In conclusion, the reported association between short term variations in air pollution and risk of death has been confirmed in a study of 968 514 deaths using a substantially different methodology than previous studies. This indicates a robustness of the association, and avoids any of the standard error problems of generalised additive models. The case-crossover approach deserves attention in the future for studies of acute health effects of pollution.

ACKNOWLEDGEMENTS

REFERENCES

1 **Maclure M**. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol* 1991;**133**:144–53.
2 **Mittleman MA**, Maclure M, Tofler GH, *et al.* Triggering of acute myocardial infarction by heavy physical exertion: protection by regular exertion. *N Engl J Med* 1993;**329**:1677–83.
3 **Neas LM**, Schwartz J, Dockery DW. A case-crossover analysis of air pollution and mortality in Philadelphia. *Environ Health Perspect* 1999;**107**:629–31.
4 **Sunyer J**, Schwartz J, Tobias A, *et al.* Patients with chronic obstructive pulmonary disease are at increased risk of death associated with urban particulate air pollution: a case-crossover analysis. *Am J Epidemiol* 2000;**151**:50–6.
5 **Levy D**, Sheppard L, Checkoway H, *et al.* A case-crossover analysis of particulate matter air pollution and out of hospital primary cardiac arrest. *Epidemiology* 2001;**12**:193–9.
6 **Lee JT**, Schwartz J. Reanalysis of the effects of air pollution on daily mortality in South Korea: a case-crossover design. *Environ Health Perspect* 1999;**107**:633–6.
7 **Schwartz J**, Dockery DW. Particulate air pollution and daily mortality in Steubenville, Ohio. *Am J Epidemiol* 1992;**135**:12–20.
8 **Schwartz J**, Dockery DW. Increased mortality in Philadelphia associated with daily air pollution concentrations. *Am Rev Respir Dis* 1992;**145**:600–4.
9 **Katsouyanni K**, Touloumi G, Spix C, *et al.* Short term effects of ambient sulphur dioxide and particulate mater on mortality in 12 European cities: results from time series data from the APHEA project. *BMJ* 1997;**314**:1658–63.
10 **Hastie T**, Tibshirani R. *Generalized additive models.* London: Chapman and Hall, 1990.
11 **Schwartz J**. Air pollution and daily mortality in Birmingham, Alabama. *Am J Epidemiol* 1993;**137**:1136–47.
12 **Katsouyanni K**, Touloumi G, Samoli E, *et al.* Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA2 project. *Epidemiology* 2001;**12**:521–31.
13 **Samet JM**, Zegar SL, Dominici F, *et al.* The National Morbidity, Mortality, and Air Pollution Study Part II: Morbidity, Mortality, and Air Pollution in the United States. Boston: Health Effects Institute, 2000;**94**:1–84.
14 **Dominici F**, McDermott A, Zeger SL, *et al.* On the use of generalized additive models in time-series studies of air pollution and health. *Am J Epidemiol* 2002;**156**:1–11.
15 **Ramsay T**, Burnett R, Krewski D. The effect of concurvity in generalized additive models linking mortality and ambient air pollution. *Epidemiology*. In press.
16 **Dominici F**. Testimony to EPA Clean Air Science Advisory Committee 18 July 2002.
17 **Schwartz J**. Testimony to EPA Clean Air Science Advisory Committee, 18 July 2002.
18 **Bateson TF**, Schwartz J. Control for seasonal variation and time trends in case-crossover studies of acute effects of environmental exposures. *Epidemiology* 1999;**10**:539–44.
19 **Bateson TF**, Schwartz J. Selection bias and confounding in case-crossover analyses of environmental time-series data. *Epidemiology* 2001;**12**:654–61.
20 **Lumley T**, Levy D. Bias in the case-crossover design: implications for studies of air pollution. *Environmetrics* 2000;**11**:689–704.
21 **Navidi W**, Weinhandl E. Risk set sampling for case-crossover designs. *Epidemiology* 2002;**13**:100–5.
22 **Schwartz J**, Zanobetti A, Bateson T. Morbidity and mortality among elderly residents in cities with daily PM measurements. In: *Revised Analyses of the National Morbidity, Mortality, and Air Pollution Study, Part II. Revised Analyses of Time-Series Studies of Air Pollution and Health. Special Report.* Boston: Health Effects Institute, 2003:25–58.
23 **Wood SN**. Modeling and smoothing parameter estimation with multiple quadratic penalties. *J R Stat Soc Ser B* 2000;**62**:413–28.
24 **Schwartz J**. Assessing confounding, effect modification, and thresholds in the association between ambient particles and daily deaths. *Environ Health Perspect* 2000;**108**:563–8.
25 **Navidi W**. Bidirectional case-crossover designs for exposures with time trends. *Biometrics* 1998;**54**:596–605.
26 **Berkey CS**, Hoaglin DC, Mossteller F, *et al.* A random-effects regression model for meta-analysis. *Stat Med* 1995;**14**:395–411.
27 **Braga ALF**, Zanobetti A, Schwartz J. The time course of weather related deaths. *Epidemiology* 2001;**12**:662–7.
28 **Curriero FC**, Heiner KS, Samet JM, *et al.* Temperature and mortality in 11 cities of the eastern United States. *Am J Epidemiol* 2002;**155**:80–7.
29 **Zanobetti A**, Schwartz J. Are diabetics more susceptible to the health effects of airborne particles? *Am J Respir Crit Care Med* 2001;**164**:831–3.