# REVIEW OF EPIDEMIOLOGICAL EVIDENCE OF HEALTH EFFECTS OF PARTICULATE AIR POLLUTION

C. Arden Pope, III[1]

Douglas W. Dockery[2,3]

Joel Schwartz[2]

For submission to:

Inhalation Toxicology

Issue reporting the Colloquium on Particulate Air Pollution

April 28, 1994

Abbreviated title: **HEALTH EFFECTS OF PARTICULATE POLLUTION**

[1]Department of Economics, Brigham Young University, 142 FOB, Provo, UT 84602.

[2]Environmental Epidemiology Program, Department of Environmental Health, Harvard School of Public Health, 665 Huntington Ave., Boston, MA 02115.

[3]The Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, 180 Longwood Ave, Boston MA 02115.

## ABSTRACT

This paper summarizes epidemiological evidence of health effects of particulate air pollution. Acute exposure to elevated levels of particulate air pollution has been associated with increased cardio-pulmonary mortality, increased hospitalization for respiratory disease, exacerbation of asthma, increased incidence and duration of respiratory symptoms, declines in lung function, and restricted activity. Small deficits in lung function, higher risk of chronic respiratory disease and symptoms, and increased mortality has also been associated with chronic exposure to respirable particulate air pollution. Health effects have been observed at levels common to many U.S. cities and at levels below current U.S. National Ambient Air Quality Standards. Although the biological mechanisms involved are poorly understood, recent epidemiological evidence supports the hypothesis that respirable particulate air pollution is an important risk factor for respiratory disease and cardio-pulmonary mortality.

While it can be argued that individual studies of daily mortality and particulate air pollution may be confounded by unmeasured or inadequately modelled covariates, taken together it is unlikely that such confounding could be consistently acting in all these studies. Daily changes in weather are expected to be associated with both air pollution and mortality, that is they are potential confounders. All the recent studies attempted to control for the effects of weather. Almost all of the studies allowed for non-linear relationships with weather factors in the analysis. The estimated particulate pollution effects are not sensitive to the inclusion of seasonal controls using approaches such as including seasonal indicator variables, including sine and cosine terms, using non-parametric smoothing techniques, or prefiltering the data. Similar particulate pollution effects have been observed in studies from both warm and cold climates, from both dry and humid locations, and from areas where particulate concentrations peak in both the summer and winter. For a comparison of these studies see Schwartz (1994e).

These studies further suggest that the observed particulate pollution effects are not confounded by associations with $SO_2$, or $O_3$. Effects observed in locations with winter peaking of particulate concentrations, when ozone concentrations are low and not a potential confounding factor are similar to effects in summer peaking areas. Most of the studies examined $SO_2$. The relationship between mortality and particulates was generally independent of $SO_2$, while the $SO_2$ relationship disappeared when particles were considered. In the Utah Valley (Pope, 1992) and Santa Clara (Fairley, 1990), $SO_2$ concentrations were low, yet the estimated $PM_{10}$ effects were similar to other effects observed in study areas with higher $SO_2$ concentrations.

**Exacerbation of asthma**

Evidence from the hospital admissions and emergency visit studies suggests that exposures to particulate air pollution may be directly associated with asthma attacks. Several investigators have considered less severe asthmatic attacks as reported by panels of asthma patients. Studies of asthmatics in the Los Angeles area (Whittemore and Korn 1980), The Netherlands (Roemer, Hoek, and Brunekreef 1993) and in Denver, Colorado (Ostro et al. 1991) reported asthmatic attacks associated with particle exposures. Bronchodilator use has also been evaluated as a measure of exacerbation of asthma in a panel of asthmatics in The Netherlands (Roemer, Hoek, Brunekreef 1993) and in Utah Valley (Pope et al. 1991). Based on the reported results of these studies the estimated percent increase in asthma attacks or use of bronchodilator associated with a 10 $\mu g/m^3$ increase in $PM_{10}$ range from 1.1% to 12% with a weighted mean of approximately 3.0%.

**Respiratory symptoms**

The use of daily diaries to record respiratory symptoms is an inexpensive method of evaluating acute changes in respiratory health status. In a commonly applied study design, panels of schoolchildren recorded the presence of specific respiratory symptoms in these daily diaries. Symptom reports are often aggregated into upper respiratory symptoms (including such symptoms as runny or stuffy nose, sinusitis, sore throat, wet cough, head cold, hay fever, and burning or red eyes) and lower respiratory symptoms (including wheezing, dry cough, phlegm, shortness of breath, and chest discomfort or pain). In addition, cough, the most frequently reported symptom, is often analyzed separately.

Studies of upper and lower respiratory symptoms had been conducted in Utah Valley (Pope et al. 1991; Pope and Dockery 1992), The Netherlands (Hoek and Brunekreef 1993, 1994), a study of six U.S. cities (Schwartz et al 1994), and Southern California (Ostro et al.

between $PM_{10}$ and grade school absences of children in Utah Valley, Utah. Lagged pollution effects of up to several weeks were observed for both restricted activity in adults and in school absences.

## CHRONIC HEALTH EFFECTS

Early cross-sectional studies suggested chronic health effects of particulate air pollution. For example, Martin (1964) reported that in the Greater London region overall annual respiratory mortality (as opposed to episodic mortality) was significantly related to smoke (or particulate pollution) levels. Holland & Reid (1965) made a cross-sectional comparison of British male postal employees in London and in smaller country towns, where levels of $SO_2$ and particulate pollution were about half those in the metropolis. Accounting for cigarette smoking levels, significant decrements of $FEV_1$ in London employees compared to those in the provinces were reported. More recent cross-sectional studies of air pollution have also reported associations between particulate pollution and respiratory symptoms, lung function and mortality rates (Table 3).

### Population-based (ecologic) mortality studies

Since the early 1970s, several studies have evaluated the mortality effects of exposure to particulate air pollution using population-based (ecologic) cross-sectional study designs (Chappie and Lave, 1982). These studies observed that on the average, mortality rates tend to be higher in cities with higher fine or sulfate particulate pollution levels than those with lower levels. Formal regression modeling techniques to evaluate cross-sectional differences in air pollution and mortality and to control for other ecologic variables was used. In an attempt to control for other risk factors, population average values for demographic variables and other factors such as smoking rates, education levels, income levels, poverty rates,

housing density and others were often included in the regression models. The basic conclusions from the population-based cross-sectional studies include: 1) Mortality rates are associated with air pollution; 2) Mortality rates are most strongly associated with fine or sulfate particulate matter; 3) An average mortality effect of 3 to 9 percent of total mortality can be estimated.

Although these population-based cross-sectional studies suggest that air pollution contributes to human mortality, the studies have been largely discounted for several reasons. One reason seems to simply be that the size of the association, if taken literally, suggests that as much as 3 to 9 percent of urban mortality in the U.S. is associated with particulate air pollution. Given that air pollution levels in the U.S. on the average are considered relatively low, such a large mortality effect seems implausible.

A more explicit limitation of the cross-sectional population-based studies is a prevailing concern that the observed association was due to confounding. Because of their ecologic design, these population-based cross sectional studies could not directly control for individual differences in cigarette smoking and other risk factors. They could only try to control for them by using population-based averages--making potential confounding a concern. Furthermore, the strength of the relationship between mortality and particulate pollution was often sensitive to model specification, socio-economic, demographic, and other risk factors included in the analysis, and the choice of study areas included in the analysis.

**Chronic differences in lung function**

There have been several recent studies that have evaluated associations between measures of lung function (FVC, $FEV_1$, PEF) and particulate pollution levels. These studies include analysis of children's lung function data from the Harvard six-city study, analysis of data from both the first and second National Health and Nutrition Examination Surveys

In all of these studies, statistically significant associations were observed between particulate air pollution and respiratory symptoms. Particulate air pollution was most consistently associated with bronchitic symptoms. The results suggest that a 10 $\mu g/m^3$ increase in $PM_{10}$ was typically associated with a 10 to 25 percent increase in bronchitis or chronic cough.

**Prospective cohort mortality studies**

Well designed prospective cohort mortality studies can provide some of the most compelling evidence about health effects because they can directly control for individual differences in age, sex, cigarette smoking, and other risk factors. Because prospective cohort studies of mortality involve collecting large amounts of information on a large number of people and following them prospectively for long periods of time, they are costly and time consuming. Currently there are only a few large prospective cohort studies that evaluate mortality effects of air pollution at levels common to U.S. urban areas.

One of these prospective cohort studies involved a 14 to 16 year prospective follow-up of 8,111 adults living in six U.S. cities (Dockery et al. 1993) TSP, $PM_{10}$, $PM_{2.5}$, $SO_4$, $H^+$, $SO_2$, $NO_2$, and $O_3$ levels were monitored. The data were then analyzed using survival analysis, including multivariate Cox proportional hazards regression modeling. Although TSP concentrations dropped over the study period, fine particulate and sulfate pollution concentrations were relatively constant. The most polluted city was Steubenville, the least polluted cities were Topeka and Portage. The Cox proportional hazards model was used to estimate adjusted mortality-rate ratios. Mortality risks were most strongly associated with cigarette smoking, but after controlling for individual differences in age, sex, cigarette smoking, body mass index, education, and occupational exposure, differences in relative mortality risks across the six cities were strongly associated with differences in pollution

significant associations were also found using fine particulate matter as the index of air pollution. Mortality due to other causes was not significantly associated with pollution levels. The association between air pollution and all-cause and cardiopulmonary mortality was observed for both men and women and for smokers and non-smokers.

## SUMMARY AND CONCLUSIONS

Numerous studies have reported that acute exposure to elevated levels of particulate air pollution is associated with changes in various human health endpoints including: 1) increased mortality, especially cardio-pulmonary mortality; 2) increased hospitalization for respiratory disease; 3) exacerbation of asthma; 4) increased incidence and duration of respiratory symptoms; 5) declines in lung function; and 6) restricted activity. Acute health effects were observed at levels common to many U.S. cities including levels well below current U.S. National Ambient Air Quality Standards. The studies generally observed that health effects increase monotonically with pollution levels, often with a near linear dose-response relationship. Furthermore these studies suggests a coherence or cascade of associations across various health endpoints (Bates 1992).

Studies have also observed human health effects associated with chronic exposure to respirable particulate air pollution. Population-based (ecologic) cross-sectional studies that evaluated spatial distributions of mortality and air pollution have observed associations between mortality and sulfate or fine particulate pollution. These population-based studies have been criticized partly because they were not able to control directly for individual differences in cigarette smoking and other risk factors. Recent studies that could adjust for individual differences in other risk factors, however, have observed that long-term exposure to respirable particulate air pollution was associated with small deficits in lung function and

higher risk of chronic respiratory disease and symptoms. Furthermore, two recent prospective cohort studies have observed increased mortality risks associated with air pollution--even after directly controlling for individual differences in age, sex, race, cigarette smoking, and other risk factors.

Although the biological linkages remain poorly understood, the results of the acute and chronic exposure studies are complementary. In all epidemiologic studies there is the concern that the observed association is due to confounding, that is, that it results from a risk factor that is correlated with both exposure and mortality but is not adequately controlled for in the study design and analysis. Important potential confounders in cross-sectional studies such as unaccounted for differences in occupational exposure or socio-economic variables, are not likely to be confounders in daily time-series studies because such factors are unlikely to change daily in correlation with air pollution levels. The fact that daily time-series studies and cross-sectional studies observe qualitatively coherent associations between respirable particulate pollution and mortality, further supports the hypothesis that this pollution is an important risk factor for respiratory disease and cardio-pulmonary mortality.

Table 1. Summary of selected acute (daily time-series) mortality studies.

| Study Area & Period | Reference | Particulate Measure | Mean[a] $PM_{10}$ ($\mu g/m^3$) | % increase in Mortality per 10 $\mu g/m^3$ increase in $PM_{10}$ (95% CI) | | |
|---|---|---|---|---|---|---|
| | | | | Total | Respiratory | Cardiovascular |
| Santa Clara, CA 1980-82, 84-86 | Fairley 1990 | Coefficient of Haze | 35 | 0.8 (0.2, 1.5) | 3.5 (1.5, 5.6) | 0.8 (0.1, 1.6) |
| Philadelphia PA 1973-80 | Schwartz & Dockery 1992 | TSP (2-day mean) | 40 | 1.2 (0.7, 1.7) | 3.3 (0.1, 6.6) | 1.7 (1.0, 2.4) |
| Utah Valley, UT 1985-89 | Pope et al 1992 | $PM_{10}$ (5-day mean) | 47 | 1.5 (0.9, 2.1) | 3.7 (0.7, 6.7) | 1.8 (0.4, 3.3) |
| Birmingham, AL 1985-88 | Schwartz 1993 | $PM_{10}$ (3-day mean) | 48 | 1.0 (0.2, 1.9) | 1.5 (-5.8, 9.4) | 1.6 (-0.5, 3.7) |
| Cincinnati, OH 1977-82 | Schwartz 1994 | TSP | 42 | 1.1 (0.5, 1.7) | 2.7 (-0.9, 6.6) | 1.4 (0.5, 2.4) |
| St. Louis, MO 1985-86 | Dockery et al 1992 | $PM_{10}$ (Prev. Day) | 28 | 1.5 (0.1, 2.9) | NA | NA |
| Kingston, TN 1985-86 | Dockery et al 1992 | $PM_{10}$ (Prev. Day) | 30 | 1.6 (-1.3, 4.6) | NA | NA |
| Detroit, MI 1973-82 | Schwartz 1991 | TSP | 48 | 1.0 (0.5, 1.6) | NA | NA |
| Steubenville, OH 1974-84 | Schwartz & Dockery 1992 | TSP | 61 | 0.7 (0.4, 1.0) | NA | NA |
| Sao Paulo, Brazil 1990-91 | Saldiva et al. 1994 | $PM_{10}$ (2-day mean) | 82 | 1.2[b] (0.7, 1.7) | NA | NA |

[a]Conversions to $PM_{10}$ assumed that: $PM_{10}$ = 0.55 * TSP and $PM_{10}$ = CoH/0.55 (for more detail on the calculations see Dockery and Pope 1994).

[b]In the Sao Paulo study mortality counts were only for the elderly (65+ years of age).

Lutz LJ. Health effects of air pollution measured by outpatient visits. J Fam Pract 1983;16:307-13.

Martin AE. Mortality and morbidity statistics and air pollution. Proc Roy Soc Med 1964;57:969-75.

Mazumdar S, Sussman N. Relationships of air pollution to health: results from the Pittsburgh study. Arch Environ Health 1983;38:17-24.

Mendelsohn R, Orcutt G. An empirical analysis of air pollution dose-response curves. J Environ Econ Manage 1979;6:85-106.

Neas LM, Dockery DW, Spengler JD, Speizer FE, Tollerud DJ. The association of ambient air pollution with twice daily peak expiratory flow measurements in children. Am Rev Respir Dis 1992;145(4):A429.

Ostro BD. The effects of air pollution on work loss and morbidity. J Environ Econ Manage 1983;10:371-82.

Ostro BD. A search for a threshold in the relationship of air pollution to mortality: a reanalysis of data on London winters. Environ Health Perspect 1984;58:397-99.

Ostro BD. Air pollution and morbidity revisited: a specification test. J Environ Econ Manage 1987;14:87-98.

Ostro BD. Associations between morbidity and alternative measures of particulate matter. Risk Analysis 1990;10:421-27.

Ostro BD. The association of air pollution and mortality: examining the case for inference. Arch Environ Health 1993; 48:336-42.

Ostro BD, Lipsett MJ, Mann JK, Krupnick A, Harrington W. Air pollution and respiratory morbidity among adults in Southern California. Am J Epidemiol 1993;137:691-700.

Ostro BD, Lipsett MJ, Wiener MB, Selner JC. Asthmatic response to airborne acid aerosols. Am J Pub Health 1991;81:694-702.

Ostro BD, Rothschild S. Air pollution and acute respiratory morbidity: an observational study of multiple pollutants. Environ Res 1989;50:238-47.

Ozkaynak H, Spengler JD. Analysis of health effects resulting from population exposures to acid precipitation precursors. Environ Health Perspect 1985;63:45-55.

Ozkaynak H, Thurston GD. Associations between 1980 U.S. mortality rates and alternative measures of airborne particle concentration. Risk Analysis 1987;7:449-61.

Ponka A. Asthma and low level air pollution in Helsinki. Arch Environ Health 1991;46:262-70.

Pope CA III. Respiratory disease associated with community air pollution and a steel mill, Utah Valley. Am J Pub Health 1989;79:623-28.

Pope CA III. Respiratory hospital admissions associated with $PM_{10}$ pollution in Utah, Salt Lake, and Cache Valleys. Arch Environ Health 1991;46:90-7.

Pope CA III, Bates DV, Raizenne ME. Health effects of particulate air pollution: time for reassessment? Am J Pub Health 1994 (in review).

Pope CA III, Dockery DW. Acute health effects of $PM_{10}$ pollution on symptomatic and asymptomatic children. Am Rev Respir Dis 1992;145:1123-28.

Pope CA III, Dockery DW, Spengler JD, Raizenne ME. Respiratory health and $PM_{10}$ pollution: a daily time series analysis. Am Rev Respir Dis 1991;144:668-74.

Pope CA III, Kanner RE. Acute effects of $PM_{10}$ pollution on pulmonary function of smokers with mild to moderate chronic obstructive pulmonary disease. Am Rev Respir Dis 1993;147;1336-40.

Pope CA III, Schwartz J, Ransom MR. Daily mortality and $PM_{10}$ pollution in Utah Valley. Arch Environ Health 1992;47:211-17.

Pope CA III, Thun MJ, Namboodiri MM, Dockery DW, Evans JS, et al. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. JAMA 1994 (in review).

Portney PR, Mullahy J. Urban air quality and chronic respiratory disease. Regional Sci Urban Econ 1990;20:407-18.

Raizenne ME, Burnett RT, Stern B, Franklin CA, Spengler JD. Acute lung function responses to ambient acid aerosol exposures in children. Environ Health Perspect 1989;79:179-85.

Ransom MR, Pope CA III. Elementary school absences and $PM_{10}$ pollution in Utah Valley. Environ Res 1992;58:204-19.

Roemer W, Hoek G, Brunekreef B. Effect of ambient winter air pollution on respiratory health of children with chronic respiratory symptoms. Am Rev Respir Dis 1993;147:118-24.

Saldiva PHN, Pope CA III, Schwartz J, Dockery DW, Lichtenfels AJ, et al. Air pollution and mortality in elderly people: a time series study in Sao Paulo, Brazil. Arch Environ Health 1994 (in review).

Samet JM, Speizer FE, Bishop Y, Spengler JD, Ferris BG Jr. The relationship between air pollution and emergency room visits in an industrial community. J Air Pollut Control Assoc 1981;31:236-40.

Schwartz J. Lung function and chronic exposure to air pollution: a cross-sectional analysis of NHANES II. Environ Res 1989;50:309-21.

Schwartz J. Particulate air pollution and daily mortality in Detroit. Environ Res 1991;56:204-13.

Schwartz J. Air pollution and daily mortality in Birmingham, AL. Am J Epidemiol 1993;137:1136-47.

Schwartz J. Air pollution and hospital admissions for the elderly in Birmingham, Al. Am J Epidemiol 1994a (in press).

Schwartz J. Particulate air pollution and chronic respiratory disease. Environ Res 1993;62:7-13.

Schwartz J. What are people dying of on high air pollution days? Environ Res 1994c;64:26-35.

Schwartz J. $PM_{10}$, ozone, and hospital admissions for the elderly in Minneapolis-St.Paul. Arch Environ Health 1994d (in press).

Schwartz J. Air pollution and daily mortality: a review and meta-analysis. Environ Res 1994e;64:36-52.

Schwartz J. Total suspended particulate matter and daily mortality in Cincinnati, Ohio. Environ Res 1994f (in press).

Schwartz J, Dockery DW. Increased mortality in Philadelphia associated with daily air pollution concentrations. Am Rev Respir Dis 1992;145:600-04.

Schwartz J, Dockery DW. Particulate air pollution and daily mortality in Steubenville, Ohio. Am J Epidemiol 1992;135:12-19.

Schwartz J, Dockery DW, Neas LM, et al. Acute effects of summer air pollution on respiratory symptom reporting in children. Am Rev Respir Dis 1994 (in press).

Schwartz J, Marcus A. Mortality and air pollution in London: a time series analysis. Am J Epidemiol 1990;131:185-94.

Schwartz J, Slater D, Larson TV, Pierson WE, Koenig JQ. Particulate air pollution and hospital emergency visits for asthma in Seattle. Am Rev Resp Dis 1993;147:826-31.

Schwartz J, Spix C, Wichmann HE, Malin E. Air pollution and acute respiratory illness in five German communities. Environ Res 1991;56:1-14.

Shumway RH, Azari AS, Pawitan Y. Modeling mortality fluctuations in Los Angeles as functions of pollution and weather effects. Environ Res 1988;45:224-41.

Shy CM. Epidemiologic evidence and the United States air quality standards. Am J Epidemiol 1979;110:661-71.

Speizer FE. Studies of acid aerosols in six cities and in a new multi-city investigation: design issues. Environ Health Perspect 1989;79:61-7.

## ABSTRACT

A cohort of 6340 non-smoking California Seventh-day Adventists (SDAs) who had resided within 5 miles of their present residence for the past 10 years have been followed since 1977 for: incidence of cancer and myocardial infarction (M.I.) through 1982, development of definite symptoms of, and increasing severity of, airway obstructive disease (AOD), chronic bronchitis, and asthma through 1987, and all natural cause mortality through 1987. Cumulative ambient concentrations of specific pollutants have been estimated for study participants from 1967 to 1987 by interpolating monthly statistics from statewide air monitoring stations to zip codes of residence and work location. Statistics include excess concentrations and exceedance frequencies above a number of cutoffs as well as mean ambient concentration and mean ambient concentration adjusted for time spent indoors. Indoor sources or nitrogen ($NO_2$), and of particulate pollution such as environmental tobacco smoke, both at home and at work, as well as occupational dusts and fumes, have been adjusted for in multivariate statistical models. Particulates included total suspended particulates (TSP), monitored from 1973-1987; inhalable particulates less than 10 microns in diameter (PM10), estimated from site/seasonal specific regressions on TSP for 1973-1987; fine particulates less than 2.5 microns in diameter estimated from airport visibility data for 1967-1987; and suspended sulfates ($SO_4$), monitored from 1977-1987. A direct measure of visibility, and gaseous pollutants--ozone, sulfur dioxide ($SO_2$), and ($NO_2$) monitored from 1973-1987 were also included in analyses.

No statistically significant associations between any of the disease outcomes studied and $NO_2$ or $SO_2$ were found in this cohort. None of the pollutants studied showed statistically

significant associations with all natural cause mortality or incidence of all malignant neoplasms in males. Statistically significant associations were observed between elevated ambient concentrations of one or more particulate pollutants and each of the other disease outcomes. In addition, ozone was significantly associated with increasing severity of asthma, and with the development of asthma in males. Multipollutant analyses indicated that none of the associations between particulate pollutants and disease outcomes were due to correlations with gaseous pollutants studied except possibly for PM2.5 and increasing severity of asthma, which could be due to a correlation with ozone. Observed associations between disease outcomes and PM2.5 or PM10 could be biased towards the null because of increased measurement error due to their indirect methods of estimation.

## INTRODUCTION

In recent years prospective studies have begun to study the association of long-term ambient concentrations of air pollutants and development of chronic disease. (Comstock, 1973; Rokaw, 1980; Detels, 1987; Dockery, 1989; EPA, 1984, 1986; Ferris, 1973,1976; Jedrychowsky, 1988; Krzyzanowski, 1990; Lebowitz, 1993; Van der Landa, 1981; Carrozzi, 1993). Most studies have dealt with respiratory disease, though Dockery et al, (1993) studied all natural cause mortality. Few studies have been able to address the effects of many different pollutants on a wide range of disease outcomes. A cohort of 6,340 non-smoking Seventh-day Adventists (SDAs) who had resided within 5 miles of their 1976 residence for the past 10 years have been followed since 1977 to ascertain the incidence of all malignant neoplasms and myocardial infarction (MI) through 1982, the development of definite symptoms of airway obstructive disease (AOD), chronic bronchitis, and asthma through 1987, and all natural cause mortality through 1986. Cumulative ambient concentrations of five specific particulate air, and three gaseous, pollutants have been estimated for study participants from 1967 to 1987 by interpolating monthly statistics from air monitoring stations statewide to zip codes of residence and work location. Previous papers (Abbey, 1989, 1991a,b; 1993a,b,c; 1994a,b,c; Euler, 1987; 1988, Hodgkin, 1984; Mills, 1991; 1993a,b;) have reported on associations between ambient concentrations of air pollutants and development of different chronic diseases in this cohort and compared the findings with other studies. The effects of occupational exposures (Greer, 1993) and environmental tobacco smoke (Robbins, 1993) have also been addressed.

For both multiple logistic and multiple linear regressions, gender, age, and education were forced into all models. Stepwise selection procedures were used to select statistically significant covariates ($p < .05$) from a large number of candidate covariates which included - years smoked prior to 1977,[1] years lived with a smoker, years worked with a smoker, years of dust exposure at work for models relating health effects to particulate pollutants, years of fume exposure at work for models relating health effects to gaseous pollutants, frequency of childhood colds, and definite symptoms of AOD before age of 16. Unless otherwise stated the term "statistically significant" will refer to the 2-tailed .05 level of statistical significance. The .05 level of statistical significance has not been used as an absolute criteria. Strong disease/pollutant associations which did not achieve the .05 level of statistical significance (possibly due to a small number of incident cases and lack of statistical power) will be described in the results below.

For incidence of MI, cancer, and all natural cause mortality Cox proportional hazards regression models were used. Further details on the epidemiological and statistical methods used are given by Abbey (1993a) and Mills (1991).

For each health outcome, multipollutant analyses compared different pollutants on the strength of association achieved by each pollutant in multivariate health effects regression models. The scale/invariant measure of strength of association used was the estimated regression coefficient for the pollutant divided by its standard error, or equivalently the computed level of statistical significance of the regression coefficient. Comparisons between pollutants were made for the mean concentration index as well as the most statistically significant of the exceedance frequency or indices. For many outcomes only one or two pollutants were associated with the

health outcome and it was clear which pollutant showed the strongest relationship. For other pollutants stepwise regression procedures were used to allow mean concentrations of the different pollutants to compete for entry in the model, if mean concentration for one of the pollutants was significantly associated with the health outcome. If this was not the case the most statistically significant of the exceedance frequency statistics for each pollutant were allowed to compete for entry in the model. In both situations gender, age, and education, and the covariates found to be significant from the single pollutant analyses were first forced into the model; then the two pollutants were allowed to compete for entry. Comparisons between TSP, ozone, $SO_4$, and PM10 were made on the entire cohort. Comparisons with each of these pollutants and PM2.5 were restricted to the subcohort living in the vicinity of airports for which PM2.5 estimates were available.

## RESULTS

Table 2 gives relative risks and 95% confidence intervals (CIs) for selected increments of ambient concentrations. Increments for either an exceedance frequency index or mean concentration were chosen to be within the range experienced by the cohort. For TSP, PM10, $SO_4$, and ozone, the increment was expressed in terms of the most statistically significant of the various indices -- mean concentration or exceedance frequencies above the different cutoffs. For PM2.5 the increment was expressed in terms of mean concentrations since it was felt that exceedance frequencies were not reliable because of discreetness due to estimating PM2.5 from visibility. Airport visibility is measured using sightings of markers at a few discreet distances from airport control towers. Some cells in Table 2 contain the entry "not done." Possible

associations between $SO_4$ and cancer incidence were not analyzed since $SO_4$ was monitored only since 1977 and it was felt that this did not allow sufficient lag time for cancer incidence. Statistical analyses to study associations of PM2.5 and visibility with respiratory cancer incidence and incidence of MI were not conducted due to an insufficient number of study participants living in the vicinity of airports for which airport visibility measures and PM2.5 estimates were available.

For the prospective data collected since 1977, no statistically significant associations were seen between any of the health outcomes studied and $SO_2$ or $NO_2$ (Abbey, 1991a; 1993a,c Mills, 1993c). This result remained true when ambient concentrations were adjusted to reflect time spent indoors and when the regression estimates of personal exposure for $NO_2$ were used. No associations were observed between long term cumulations of any of the ambient pollutants studied and incidence of all malignant neoplasms in males, or all natural cause mortality. Relative risks and confidence intervals for these two outcomes as associated with elevated TSP and ozone have been reported previously. (Abbey, 1991a; 1993c, Mills 1993c) Relative risks for these two health outcomes for all pollutants studied were close to, or less than, one. The null relative risk of one was well within all 95% confidence intervals. Incidence of MI was significantly associated only with exceedance frequencies above 6 $\mu g/m^3$ of $SO_4$ (Abbey, 1993b). No associations were observed between visibility and health outcomes except for mean visibility and increasing severity of asthma symptoms and mean visibility and incidence of all malignant neoplasms in females. These observed associations with visibility, however, were not as

statistically significant as those with PM2.5 and were likely the result of the correlation between PM2.5 and visibility.

Table 3 gives the cutoffs for which exceedance frequencies show statistical significance and the results of the multipollutant analyses. Pollutants in parenthesis in Table 3 indicated that the observed association with a health outcome could be due instead to an association of the pollutant in parenthesis and the health outcome, as the pollutant in parenthesis showed a stronger association.

TSP showed the strongest associations of any of the pollutants with development of AOD, and chronic bronchitis, as well as increasing severity of symptoms for AOD, and chronic bronchitis. For asthma, TSP, ozone, and $SO_4$, were close competitors, (see ozone and $SO_4$ sections below). TSP was statistically significantly associated with incidence of all malignant neoplasms in females, but not males. When analyses were restricted to the airport sub-cohort for which estimates of PM2.5 were available, TSP remained statistically significantly associated with incidence of all malignant neoplasms in females but the association was not as strong as for PM2.5. From Table 3 it can been seen that ambient concentrations of TSP in excess of 60 $\mu g/m^3$ were associated with increasing severity of AOD symptoms. Ambient concentrations of TSP in excess of 100 $\mu g/m^3$ were associated with many of the health outcomes.

Ozone was highly significantly associated with development of asthma in males but not females. This relationship does not appear to be a surrogate relationship. The relative risk for a 1 pphm increase in mean concentration of ozone was 3.12, (95% CI: 1.61, 5.85), (Greer, 1993). In non-gender specific analyses exceedance above 10 pphm of ozone showed a trend

association ($p = .056$) with development of asthma. Ozone was more strongly associated with development of asthma than TSP when mean concentrations of the two pollutants were allowed to compete for entry in the models but was less strongly related when the most significant exceedance frequency statistic of each pollutant was used, i.e., TSP in excess of 200 $\mu g/m^3$ showed a stronger association with development of asthma than ozone in excess of 10 pphm. Mean concentration of ozone and exceedance frequencies for the two lowest cutoffs, 10 pphm and 12 pphm, show statistical significance when associated with increasing severity of asthma. For increasing severity of asthma, stepwise linear regression showed that mean concentration of ozone enters in preference to mean concentration of TSP. However, in another stepwise multiple linear regression; the F to enter for exceedance frequencies in excess of 200 $\mu g/m^3$ TSP was tied with the F to enter for exceedance frequencies in excess of 10 pphm of ozone. Ozone showed a strong trend association with respiratory cancer incidence. Due to the small number of incident cases, only 17, this association was not statistically significant. Future research will investigate lower cutoffs of ozone as well as 8-hour averages.

$SO_4$ was statistically significantly associated with development of asthma and this association was stronger than that for TSP or ozone when the covariate, "possible symptoms in 1977" was not included in the model. Analyses based on the entire respiratory symptoms cohort indicated that the relationship between $SO_4$ and increasing severity of AOD symptoms may instead be due to TSP and the correlation between $SO_4$ and TSP. When analyses were restricted to the airport sub-cohort, however, $SO_4$ was more strongly associated with increasing severity of AOD symptoms than any of the other pollutants including TSP.

caution. The primary purpose of Table 3 is to show that statistically significant associations between disease outcomes and particulate pollutants have been observed in this cohort at ambient levels in excess of ambient concentrations which frequently occur in urban areas of California, and that these associations appear not to be surrogate relationships solely due to correlations with gaseous pollutants, which were monitored during the time period of this study.

Estimated ambient concentrations of pollutants for study participants were averaged over different time periods before being used in analyses. The stepwise selection procedures used in forming the regression models were allowed to select the most statistically significant time period. The previous papers from this study show which time period for each pollutant was most significantly related to each health outcome. Except for $SO_4$ and PM2.5 the time period 1973-1977 was used as a surrogate for the time period 1966-1977 as many more monitoring stations were available for the latter time period. $SO_4$ was monitored only since 1977. PM2.5 was estimated from airport visibility data since 1966. Sensitivity analyses showed that results using either time period agreed closely. Cox proportional hazards regression models, which were used for incidence and mortality, where date of event was available, used an average of the pollutant from 1973 through data of event or risk set (time dependent models) as well as the average from 1973 through March 1977, (fixed time models). Sensitivity analyses indicated concurrence between the two types of models.

Our respiratory symptoms complex of chronic bronchitis included chronic productive cough (with sputum) as well as cough only. Development of definite symptoms of chronic bronchitis was statistically significantly associated with elevated ambient concentrations of TSP,

cities in years 8-16 of follow-up. We only had 10 years of mortality follow-up available for present analyses on our cohort. This may explain the discrepancy.

We observed an elevated risk of all malignant neoplasms in females but not in males as associated with TSP and PM2.5. A similar result has been observed in an epidemiological study in Israel (Biger, personal communication, 1991). Experiments in rats have shown that particulate matter in diesel exhaust fumes have a greater impact on lung tumor formation in females than males especially at the highest concentration levels (Mauderly, 1987). Fine particulates often contain the most toxic compounds (e.g., trace metals, acid sulfates, organics, etc.) (Özkaynak, 1987). Females have been shown to have a greater deposition fraction than males of fine particles in the lungs (Kim, 1994).

The increased risk of respiratory cancer associated with elevated ozone in our cohort is consistent with some animal studies. Dillon (1992) showed that ozone was mutagenic in Salmonella. Ozone has been shown to be carcinogenic in mice though not in rats. (Hassett, 1985; Last, 1987; NTP Technical Report, 1993). Our findings of elevated risks of asthma and respiratory cancer as associated with ozone is consistent with a significantly increased standardized incidence ratio of lung cancer observed in asthmatics in Finland (Vesterinin, 1993), and a very slightly elevated, but not statistically significant standardized morbidity ratio found for respiratory tract cancer in asthmatics in Sweden (Källén, 1993).

Ozone was significantly associated with development of asthma in males but not in females. T-test between males and females in our cohort showed that males were outdoors significantly more than females during the high ozone season of June through September (Greer,

1993). Ozone is highly volatile and not stable in the indoor environment. This may explain the gender discrepancy with respect to development of asthma.

We found ozone to be associated with asthma and asthmatic symptoms but not chronic bronchitis. This pattern of relationships has been observed in other studies (von Mutius et al., 1992; Viegi et al., 1991; Krzyzanowski et al, 1990; Lippmann, 1989). Chronic bronchitis and chronic obstructive pulmonary disease (COPD) appear to be associated much more with reducing type (PM-$SO_x$) atmospheres (Sunyer et al., 1993; Lebowitz, 1983; Ware, 1980; ATS, 1978; Colley & Holland, 1967). Our findings that TSP and PM10 are more strongly associated with development of and an increasing severity of symptoms of overall AOD and chronic bronchitis than other pollutants are consistent with these findings. However, in our study $SO_4$ was strongly associated with development of asthma as well as increasing severity of AOD.

Lung function parameters such as post bronchodilator response as well as lability indices from home peak flow diary data have recently been collected on 1,500 members of our cohort. An additional 10 years of cancer incidence data is being collected on the entire cohort as well as an additional six years of mortality. Ambient air pollution indices are being updated on the cohort through 1993. These updated indices will include monitored PM10, available on a statewide basis since 1987. New indices for ozone are being computed which include 8 hour averages and exceedance frequencies and excess concentrations for two new lower cutoffs, 6 and 8 pphm. Analyses of these new data may provide answers to some of the questions raised from current findings.

Abbey, D.E., Colome, S.D., Mills, P.K., Burchette, R, Beeson, W.L., Tian, Y. (1993c), "Chronic Disease Associated With Long Term Concentrations of Nitrogen Dioxide," *Journal of Exposure Analysis and Environmental Epidemiology*, 3(2), 181-202.

Abbey, D.E., Hwang, B.L., Burchette, R.J. (1994a), "Long-Term Ambient Concentrations of PM10 and Airway Obstructive Disease," *Archives Environmental Health*, In press

Abbey, D.E., Ostro, B.E., Petersen, F., Burchette, R.J., VanCurren, T., Fraser, G. (1994b), "Chronic Respiratory Disease Associated With Long-Term Ambient Concentrations of Fine Particulates Less than 2.5 Microns in Diameter (PM2.5) and Other Air Pollutants," *Archives Environmental Health*, Submitted for publication.

Abbey, D.E. (March, 1994c) "Incidence of Respiratory Symptoms and Chronic Disease in a Non-Smoking Population as a Function of Long-Term Cumulative Exposure to Ambient Air Pollutants (AHSMOG Study)," *CARB Final Report*, Contract Number A933-160.

ATS. (1978) "Statement on the Health Effects of Air Pollution," *American Thoracic Society*, New York.

Carrozzi, L. (1993), "Longitudinal Analysis of Respiratory Symptoms in a General Population Sample of Northern Italy," *European Respiratory Journal*, 6(S17), 408s.

Colley, J.R.T., Holland, W.W. (1967), "Social and Environmental Factors in Respiratory Diseases," *Archives of Environmental Health*, 14, 157.

Comstock, G.W., Stone, R.W., Sakai, Y., Matsuya, T., Tonascia, J.A. (1973), "Respiratory Findings and Urban Living," *Archives of Environmental Health*, 27, 143-150.

to Ambient Levels of Total Suspended Particulates and Sulfur Dioxide in California Seventh-Day Adventist Residents," *Archives of Environmental Health*, 42(4), 213-222.

Euler, G.L., Abbey, D.E., Hodgkin, J.E., Magie, A.R. (July/August, 1988), "Chronic Obstructive Pulmonary Disease Symptom Effects of Long-Term Cumulative Exposure to Ambient Levels of Total Oxidants and Nitrogen Dioxide in California Seventh-Day Adventist Residents," *Archives of Environmental Health*, 43(4), 279-285.

Ferris, B.G., Higgins, I.T.T., Higgins, M.W., Peters, J.M. (1973), "Sulfur Oxides and Suspended Particulates: Possible Effects of Chronic Exposure," *Archives of Environmental Health*, 27, 179.

Ferris, B.G., Chen, H., Pulea, S., Murphy, R.L. (1976), "Chronic Non-Specific Respiratory Disease in Berlin, New Hampshire, 1967-73," *American Review of Respiratory Disease*, 113, 475.

Greer, J.R., Abbey, D.E., Burchette, R.J. (1993), "Asthma Related to Occupational and Ambient Air Pollutants in Nonsmokers," *Journal of Occupational and Environmental Medicine*, 35(9), 909-915.

Hassett, C., Mustafa, M.G., Coulson, W.F., and Elashoff, R.M. (1985), "Murine Lung Carcinogenesis Following Exposure to Ambient Ozone Concentrations," *Journal of National Cancer Institute*, 75, 771-777.

Hodgkin, J.E., Abbey, D.E., Euler, G., Magie, A.R. (December, 1984), "COPD Prevalence in Nonsmokers in High and Low Photochemical Air Pollution Areas," *Chest*, 86, 830-838.

Jedrychowsky, W., Krzyzanowski, M. (1988), "Ventilatory Lung Function and Chronic Chest Symptoms Among the Inhabitants of Urban Areas With Various Levels of Acid Aerosols: Prospective Study in Cracow," *Environmental Health Perspectives*, 79, 101-107.

Källén, B., Gunnarskog, J., Conradson, T-B. (1993), "Cancer Risk in Asthmatic Subjects Selected from Hospital Discharge Registry," *European Respiratory Journal*, 6, 694-697.

Kim, C.S., Hu, S.C., Dewitt, P., and Gerrity, T.R. (January 24-25, 1994), "Assessment of Local Deposition of Inhaled Particles in Human Lungs," *Colloquium on Particulate Air Pollution*, University of California, Irvine. Session P2.6 Presentation.

Krzyzanowski, M., Camilli, A.E., Lebowitz, M.D. (1990), "Relationships Between Pulmonary Function and Changes in Chronic Respiratory Symptoms - Comparison of Tucson and Cracow Longitudinal Studies," *Chest*, 98, 62-70.

Last, J.A., Warren, D.L., Pecquet-Goad, E., Witschi, H. (1987), "Modification by Ozone of Lung Tumour Development in Mice," *Journal of National Cancer Institute*, 78, 149-154.

Lebowitz, M.D. (1983), "Utilization of Data from Human Population Studies for Setting Air Quality Standards: Evaluation of Issues," *Environmental Health Perspectives*, 52, 193-205.

Lebowtiz, M.D., (1993), "Pulmonary Responses to Multipollutant Airborne Particulate Matter and Other Contaminants," in *Prevention of Respiratory Diseases*, eds. A. Hirah, et al. New York: Marcell Dekker, pp. 209-223.

Lippmann, M. (1989), "Health Effects of Ozone. A Critical Review," *Journal of Air Pollution Control Association*, 39, 672-95.

Özkaynak, H., Thurston, G.D. (1987). "Associations Between 1980 U.S. Mortality Rates and Alternative Measures of Airborne Particle Concentration," *Risk Analysis*, 7, 449-60.

Robbins, A.S., Abbey, D.E. (1993), "Passive Smoking and Chronic Respiratory Disease Symptoms in Non-Smoking Adults," *International Journal of Epidemiology*, 22, 809-817.

Rokaw, S.N., Detels, R., Coulson, A.H., Sayre, J.W., Taskin, D.P., Allwright, S.S., Massey, F.J. (1980), "The UCLA Population Studies of Chronic Obstructive Pulmonary Disease. "Comparison of pulmonary function in three communities exposed to photo-chemical oxidants, multiple primary pollutants or minimal pollutants," *Chest*, 78, 252-262.

Sawicki, F., Lawrence, P.S. (1977), "Chronic Non-Specific Respiratory Disease in the City of Cracow," *National Institute of Hygiene*, Warsaw, Poland.

Shy, C.M., Creason, J.P., Pearlman, M.E., McClain, K.E., Benson, F.B., Young, M.M. (1970), "The Chattanooga School Study: Effects of Community Exposure to Nitrogen Dioxide. II. Incidence of Acuter Respiratory Disease," *Journal of Air Pollution Control Association*, 20, 582-288.

Sunyer, J., Saez, M., Murillo, C., et al. (1993), "Air Pollution and Emergency Room Admissions for COPD: A five Year Study," *American Journal of Epidemiology*, 1273, 701-05.

Van der Landa, R. (1981), "Longterm Exposure to Air Pollution and Declines in Lung Function," *Chest*, 80s, 23-26.

Vesterinen, E., Pukkala, E., Timonen, T., Aromaa, A. (1993), "Cancer Incidence Among 78,000 Asthmatic Patients, *International Journal of Epidemiology*, 22, 976-982.

Viegi, G., Paoletti, P., Carrozzi, L., Guintini, C., Lebowitz, M.D. (1991), "Prevalence Rates of Respiratory Symptoms in Italian General Population Samples Exposed to Different Levels of Air Pollution, *Environmental Health Perspectives*, 94, 95-99.

von Mutius, E., Fritzsch, C., Weiland, S.K., et al. (1992), Prevalence of Asthma and Allergic Disorders Among Children in United Germany: A Descriptive Comparison," *British Medical Journal*, 305, 1395-1399.

Ware, J. (1980), "Assessment of the Health Effects of $SO_x$ and PM: Analysis of the Exposure-Response Relationship," *U.S. EPA*, Research Triangle Park, NC.

Winer, A.M., Lurman, R.W., Coyner, L.A., Colome, S.D., Poe, M.P. (1989), "Characterization of Air Pollutant Exposures in the California South Coast Air Basin: Applications of a New Regional Human Exposure (REHEX) Model," Final Report. Contract No. TSA 106-01-88. *Statewide Air Pollution Research Center.* University of California, Riverside, California.

Footnote (page #8)

[1] Many SDAs join the church later in life and have smoked before becoming an SDA; any individuals who reported smoking since 1977 were excluded.

**Table 1 Pairwise Correlations of Estimated Mean 1977-1987 Concentrations of Ambient Pollutants for Members of the AHSMOG Respiratory Symptoms Cohort (n = 3914)[a]**

| | TSP | PM10[b] | PM2.5[c] | $SO_4$ | OZ | $SO_2$ | $NO_2$ |
|---|---|---|---|---|---|---|---|
| TSP | 1 | .95 | .86 | .69 | .72 | .61 | .46 |
| PM10[b] | . . . | 1 | .89 | .72 | .76 | .64 | .52 |
| PM2.5[c] | . . . | . . . | 1 | .30 | .62 | .47 | .25 |
| $SO_4$ | . . . | . . . | . . . | 1 | .57 | .60 | .63 |
| OZ | . . . | . . . | . . . | . . . | 1 | .38 | .36 |
| $SO_2$ | . . . | . . . | . . . | . . . | . . | 1 | .85 |
| $NO_2$ | . . . | . . . | . . . | . . . | . . . | . . . | 1 |

[a] Correlations between PM2.5 and other ambient concentrations are computed only for 1977-1986 for individuals living in the vicinity of nine California airports, n = 1868. Correlations of other pollutants are computed for the entire cohort, n = 3914, April 1977 - April, 1987.

[b] PM10 was indirectly estimated from TSP using site/seasonal regression equations.

[c] PM2.5 was indirectly estimated from airport visibility data.

## Introduction

Since the late 1970s, rates of asthma morbidity and mortality have increased strikingly in the United States, with African-Americans being disproportionately affected (Sly 1988; Evans et al. 1987; Weiss et al. 1990). Although asthma prevalence is only somewhat greater among African Americans than among whites, rates of morbidity and mortality are markedly higher. Part of the racial differences in rates are likely due to conditions associated with poverty (Wissow 1988). Various social and environmental factors, including exposure to indoor and outdoor pollutants and allergens, have been postulated as explanations of these asthma trends (Weiss, 1993). To date, however, little research has been conducted to determine the effects of environmental factors, such as air pollution, on asthma among African Americans.

Epidemiologic evidence indicates that ozone and particulate matter are associated with exacerbations of asthma (Schwartz et al., 1993; Thurston et al., 1992; Pope, 1991). Most epidemiologic studies have relied on measurement of airborne particles other than PM10 (particulate matter less than 10 microns in diameter), the current metric for the National Ambient Air Quality Standard for particulate matter in the United States. In most areas, particulate matter is measured only every sixth day. The lack of daily data on ambient levels of PM10 has impeded epidemiologic research, while the complex and heterogeneous nature of particulate matter has discouraged clinical efforts to study its effects in the laboratory. Although results of several epidemiologic studies suggest that ambient ozone is associated with exacerbations of asthma, the results from controlled exposure studies are mixed, especially at low effective doses (Kreit et al., 1989; Koenig et al., 1985; Linn et al., 1978).

In this paper, we report results of a pilot study of the respiratory health of a sentinel group of asthmatic children in relation to several ambient pollutants in Southern California. The results suggest that relatively low ambient levels of both PM10 and ozone are associated with shortness of breath among this population of African-American children.

## Data and Methodology

levels of pollen from grasses, trees and weeds, and specific molds, such as alternaria and cladosporium). Levels of pollens and molds exhibited very distinct peaks and were not correlated with either PM10 or ozone concentrations. Temperature, humidity, and levels of pollens and molds were considered both as continuous variables and as binary variables indicating extremes. Specifically, we examined the effect of a given variable being in the top 25, 10, and 5 percent of all such measurements. As indicated in Figure 2, the levels of molds (and pollen, not displayed) peaked at the end of the study period. In general, there appeared to be only minimal confounding among the pollutants, pollens, and molds. Table 2 details the distribution of the levels of air pollutants, pollens, molds, and three of the weather variables.

Statistical Methods

In a panel design study each person can serve as his or her own control. The data set therefore will provide substantial degrees of freedom and a reduced potential for confounding and covariation among explanatory variables.

Three different levels of analysis were conducted for each of the six binary response variables indicating daily symptoms: the prevalence (i.e., the likelihood of a symptom on a given day) of cough, shortness of breath and wheeze, and the likelihood of an incident attack, for any of these three asthma symptoms, that might extend over several days. In the first analysis, qualitative in nature, the group rate of symptoms (e.g., the proportion of the sample reporting cough on a given day) was analyzed. This was undertaken in order to detect general patterns in the data, to determine the form of the dose-response relationship, to narrow down the list of candidate pollution and meteorologic variables, to examine patterns of residuals for autocorrelation, and to search for influential data points.

Next, all of the individual observations were pooled and examined by logistic regressions based on the generalized estimating equations (GEE) proposed by Liang and Zeger (1986). This method generates robust estimators regardless of the specification of the covariance matrix and corrects for the repeated measure design of the study. Thus, the standard error of the regression estimate is adjusted for the fact that responses

the sensitivity of the results to alternative regression specifications, we reported the odds ratios and confidence intervals as additional variables were added to the model. Potential interactions among the included variables were also examined.

RESULTS

1. Pooled analysis

With the exception of the reporting of a symptom or a cold on the previous day, most of the variables were not associated with the reporting of asthma symptoms. Of the pollutants, PM10 and ozone were associated with shortness of breath but not with any other symptom. Accordingly, shortness of breath was the only symptom examined in subsequent models. The generalized estimating equations included a one-period autocorrelation correction that effectively reduced the serial correlation of the residuals. Table 3 shows the association of shortness of breath with PM10 and ozone concentrations as additional covariates are added to the model specification. As explanatory variables, the basic model includes a binary variable indicating the presence of shortness of breath on the previous day, a binary variable for respiratory infection on the previous day, and one air pollution variable, either PM10 or ozone. Shortness of breath had a statistically significant association with both ozone (OR = 1.40, 95%CI = 1.14 - 1.72 evaluated at the mean ozone level) and PM10 (OR = 1.58, 95%CI = 1.05 - 2.38 evaluated at the mean PM10 level). Having a respiratory infection on the previous day was statistically associated with shortness of breath ($p < 0.05$). Transforming the pollution variables into log terms only slightly improved the association of shortness of breath with PM10 and had no impact on its association with ozone. Contemporaneous exposures of ozone and PM10 were more associated with symptoms than were the lagged versions of these variables. Therefore, in subsequent analyses, we used untransformed same-day exposures of pollution in the regression specifications. When children were assigned exposures based on either the average of the three fixed-site monitors or the monitor closest to their residence, the significance of the associations with PM10 and ozone did not change.

When continuous values of temperature, pollens and molds were added to the model, only temperature was associated with shortness of breath (p

- .10). This association was unchanged when temperature extremes (i.e., days with the highest 10% maximum temperatures, greater than 82° F) were added to the model. Although the binary variables representing the extreme values for molds were not statistically significant, a binary variable indicating the 5% of days with highest levels of molds (i.e., days on which more than 2,000 spores per cubic meter were recorded) was kept in subsequent models because the t-statistic was close to one. The inclusion of these variables did not affect either the magnitude or significance of the coefficient for PM10. The statistical significance of ozone dropped slightly ($p < 0.05$) probably because of the covariation between ozone and temperature. When we added the variable indicating a previous day with a cold back into the model, the pollution results were not altered, although in the specification with ozone, temperature was now more significant ($p < 0.05$). There did not appear to be any interactions between pollutants or between pollution and either temperature or molds. Among the individual demographic variables, the frequency of symptom reporting did not vary on the basis of sex, but frequency did increase as age and family income increased.

For the subsample of the population with moderate or severe asthma, shortness of breath was associated with both PM10 and ozone but the magnitude of the effect was similar to the group as a whole. Finally, the beginning of an asthma episode was associated with ozone ($p < 0.05$) (OR = 1.60, 95%CI = 1.01-2.58) but not PM10 ($p > 0.10$) (OR = 1.06,95%CI = 0.61-1.84).

2. Individual-level analysis

As expected, many children were excluded from this part of the analysis because of low symptom rates or too few days of response, leaving only 41 subjects. Again, no association was found between any time-varying factors, including air pollution, and either cough or wheeze. However, shortness of breath was associated with both PM10 ($p < 0.01$) (OR = 2.42, 95%CI = 1.76-4.68) and ozone ($p = 0.06$) (OR= 1.51, 95%CI = 0.98-2.32). Also, the beginning of an incident attack (or episode) of shortness of breath was associated with ozone, but not PM10. Finally, to compare the models, we reran the pooled analysis using the same subsample of 41 subjects and found both ozone and PM10 to be associated with

approach, in which each individual served as his or her own control, no other factor is likely to be correlated with both outdoor air pollution and shortness of breath. Although this approach is based on a more sensitive subgroup of the sample (those reporting enough symptom days or total days so that the individual logistic regressions could converge) it suggests that there are some children who are especially sensitive to air pollution, especially PM10.

The biologic validity of the consistent statistical relationship between PM10, ozone, and reported shortness of breath would be strengthened by finding similar associations with wheezing. However, individual perceptions of shortness of breath may occur in relation to a sensation of chest tightness with or without audible wheezing. This seems to be the case with our study population: reporting a shortness or breath and chest tightness were strongly associated ($r = 0.42$; $p < 0.0001$).

In general, the impact of air pollution on this population of African-American children in Los Angeles with asthma appears significant. At the mean levels of pollution, we detected a 50 percent increase in the reporting of shortness of breath. For a 10 $\mu g/m^3$ change in PM10, this amounts to about a 9 percent increase in symptoms.

782.

Whittemore AS, Korn EL (1980) Asthma and air pollution in the Los Angeles area. Am J Public Health 70:687-696.

Zeger S, Liang K (1979) Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42:121-130.

Table 1. Demographic and Socioeconomic Characteristics of Children who Completed Diaries (n = 83).

| Characteristics | Categories | Percent (%) |
|---|---|---|
| Age Distribution (years) | 7-8 | 33.7 |
| | 9-10 | 37.3 |
| | 11-12 | 28.9 |
| Sex | Female | 32.5 |
| | Male | 67.5 |
| Annual Household Income | < $10,000 | 20.8 |
| | $10-25,000 | 24.7 |
| | $25-40,000 | 24.7 |
| | $40-70,000 | 23.4 |
| | > $70,000 | 6.5 |
| Asthma Severity | Mild | 28.9 |
| | Moderate | 53.0 |
| | Severe | 18.1 |
| Use Inhaled Steroids | Yes | 27.7 |
| | No | 72.3 |
| Use any Anti-inflammatory Medicine | Yes | 45.8 |
| | No | 54.2 |
| Recruitment Site | Kaiser West L.A. | 18.1 |
| | Kaiser Inglewood | 28.9 |
| | King-Drew Med. Center | 9.6 |
| | Children's Hospital, L.A. | 8.4 |
| | Asthma Camps | 34.9 |

# Excess Mortality and Exposure to $PM_{10}$ Air Pollution: Additional Analyses from Utah

Joseph L. Lyon, M.D., M.P.H.

Motomi Mori, Ph.D.

Renlu Gao, M.Stat

Department of Family and Preventive Medicine, University of Utah School of Medicine, Salt Lake City, Utah

Address Correspondence to:

Joseph L. Lyon, M.D., M.P.H.

Deparatment of Family and Preventive Medicine

University of Utah School of Medicine

50 North Medical Drive, Room 1C26

Salt Lake City, UT 84132

ABSTRACT

We examined the association between exposure to $PM_{10}$ air pollution and daily mortality in Utah County, Utah for 1985-92. We confirmed the previous finding that exposure to fine particulate air pollution (particulate diameter of ≤ 10 μm) in amounts of 50+ $\mu g/m^3$ increased daily mortality by 3%. The potential importance of this observation led us to test the association more rigorously by assessing rate ratios (RR) of $PM_{10}$ for year, season, and location at time of death. For individual years there was no statistically significant association between increased mortality and exposure to $PM_{10}$ air pollution. The strongest mortality effect was seen in the spring, not the winter. The largest numeric contribution to excess mortality was from individuals age 75+ dying in hospital, and the largest RR was for individuals ages 15-59 dying at home primarily of cancer. These findings do not support a causal association between exposure to $PM_{10}$ air pollution and daily mortality.

INTRODUCTION

Daily exposure to fine particulate air pollution of average diameter 10 μ or less ($PM_{10}$) has been associated with increased mortality within a few days of exposure after controlling for relevant weather variables (Schwartz, 1991; Schwartz and Dockery, 1992a; Schwartz and Dockery, 1992b; Schwartz, 1993; and Pope et al., 1992). The excess mortality was greatest among those dying from cardiopulmonary diseases, and was present at levels of air pollution below the current National Ambient Air Quality Standard. These studies have assumed uniform exposure to air pollution within a city or county, regardless of where the death occurred. However, many people die in hospitals or nursing homes rather than at home, and it is questionable if these individuals are uniformly exposed to the same level of $PM_{10}$. None of the previous studies have determined whether the association differed by location at time of death, or was uniform across seasons and the years studied. The failure to find a consistent mortality effect across time or by location would suggest that the excess mortality attributed to air pollution was the result of confounding by other factors such as weather, or effect modification by an unknown factor.

Because of the potential impact of this association on the public health, this study tested the association between exposure to $PM_{10}$ air pollution and excess mortality more rigorously than previous studies. The study objectives were: (1) to examine the excess mortality associated with exposure to fine particulate air pollution by age groups and location at time of death; (2) to examine the causes of death associated with $PM_{10}$ exposure by age

group, location at time of death, and by year and season; (3) and to examine the annual and seasonal variation of the $PM_{10}$ effect on daily mortality.

County always occurred during the winter, and were always associated with low temperature and high humidity.

Age groups are categorized as <1, 1-14, 15-59, 60-74, and 75+. Location at time of death was defined as a hospital death if it occurred in one of the four hospitals serving the county. A nursing home death if it occurred in one of the nursing homes located in the county, and "home" if the death occurred outside of a nursing home or hospital. Based on the death certificate, 98.2% of deaths outside a hospital or nursing home occurred at home.

Since the number of deaths ranged from 0 to 10 with the mean of 2.52 and the variance of 2.68, the data were analyzed using a Poisson regression model. The generalized estimating equation (GEE) approach, previously used in the analysis of daily mortality and air pollution (e.g., Schwartz et al., 1992), was not used since there is very little correlation among death counts. For example, Lag 1 to 60 autocorrelations were all in the range of -0.05 and 0.05. Consequently, daily death counts were assumed to be independent and were analyzed using the standard Poisson regression program available in EGRET (SERC 1991). Daily minimum temperature was entered as a continuous variable, while $PM_{10}$ was entered as a dichotomous variable using 50 $\mu g/m^3$ as a cutoff (i.e., 1 if above 50 and 0 otherwise). Justification for this dichotomy comes from the dose-response relationship (see Figure 1). A linear time trend was included as a continuous variable to adjust for an increase or decrease in background mortality rates over the years.

Variations in $PM_{10}$ effect by season are present in Table 4. The largest number of excess deaths occurred during spring (April, May, June), followed by winter (January, February, March).

Causes of death were divided into three broad categories (respiratory, cardiovascular and other), and RRs of $PM_{10} \geq 50\ \mu g/m^3$ were calculated for each cause as well as by year of death and season of death (see Tables 3 and 4). Overall the RRs for cardiovascular deaths were similar to the RRs for pulmonary deaths; however, this varied within each year. The RR for respiratory deaths was highest in 1986 and 1987 when $PM_{10}$ levels were quite low, was below 1 in 1988, a year with high $PM_{10}$, and was 1.0 in 1990, the year when $PM_{10}$ levels were the highest. The $PM_{10}$ effect for respiratory causes of death was strongest in the spring with an RR of 1.47, while that for cardiovascular causes was greatest in the winter (RR=1.22). When the data were examined by season and location at time of death, the strongest $PM_{10}$ effect was found for deaths at locations other than hospital or nursing home (RR=1.33 for spring, and 1.21 for winter), followed by deaths occurring in hospital (RR=1.09 for spring and 1.05 for winter). (See Table 5.)

DISCUSSION AND CONCLUSION

This study examined the association between $PM_{10}$ and daily mortality in greater detail than previously published studies to assess the causality of the association. The lack of a consistent association for any calendar year, and in the lower RR for the winter season calls into question the causality of the association.

This study confirmed the findings previously reported by Pope et al. (1992) of excess mortality in Utah County associated with fine particulate air pollution. The coefficients obtained from for this study are in agreement with those of the previous study, but the standard errors are slightly bigger. Problems in the calculation of standard error in such studies have been reported previously (Mori et al., 1994).

No previous study has examined the $PM_{10}$ effect by the location at time of death. Our findings suggest an effect for those age <1-59 who died at home (n=366), but these deaths were principally due to "other" than respiratory causes such as sudden infant death syndrome and cancer. The largest overall contribution to the excess deaths was from those age 75+ who died in the hospital (n=1585). There is no information on $PM_{10}$ levels in the hospital compared to $PM_{10}$ levels at home or outdoors. Hospital ventilation systems usually draw in outside air, condition it, and then pass it to the wards. The airflow is generally from patient rooms to the corridors. No information is available for $PM_{10}$ levels in intensive care units, patients on respirators, in isolettes or receiving

supplemental oxygen, making it difficult to draw conclusions regarding the effect of outside $PM_{10}$ levels on hospital mortality.

Excess mortality was not present every year, and excess deaths from respiratory disease, the cause believed to be most strongly influenced by $PM_{10}$ exposure, followed a pattern independent of overall excess mortality by year. Further, this study found no effect of $PM_{10}$ exposure at levels of less than 50 $\mu g/m^3$ contrary to findings from previous studies (Schwartz, 1993).

The lack of an association in certain years raises a question about the causal nature of $PM_{10}$ air pollution and subsequent mortality. The lack of association (RR=1.0) in 1987 might be explained by a "threshold effect" where so many days of high $PM_{10}$ exposure is necessary before a mortality effect occurs. $PM_{10}$ levels in 1987 were the lowest of all the years studied. However, this explanation cannot account for the lack of association in 1991 (RR=1.0) and 1992 (RR=1.01) when $PM_{10}$ levels were as high as any recorded in 1985, the year with the highest RR (RR=1.20). Nor can this explanation account for the RR in 1990 (RR=1.20), a year that had the second lowest $PM_{10}$ levels after 1987.

The RRs for season are also perplexing. The highest levels of $PM_{10}$ were reported in the winter (January and February), but the strongest effect occurs during the spring (RR=1.13), followed by winter (RR=1.08). There is no effect seen in the six months from July to December (RR=0.90 and 0.97). When RRs were calculated for each month the strongest effect occurred in May.

The cause of death was assigned by the State of Utah, and no effort was made to check its accuracy or recode the death

BIBLIOGRAPHY

Mori M., Lyon J. L., Carter R. A., Self, S. G. 1994. Evaluation of statistical method for assessing association between daily mortality and air pollution. Submitted.

Pope C. A., Schwartz J., Ransom, M. R. 1992. Daily mortality and $PM_{10}$ pollution in Utah Valley. Arch Environ Health. 47:211-217.

Pope C. A., Dockery, D.W., Spengler, J.D., Raizenne, M.E. 1991. Respiratory health and $PM_{10}$ pollution: a daily time series. Am Rev Respir Dis. 144:668-74.

Statistics and Epidemiology Research Corporation (SERC). EGRET. Seattle, Washington, 1991.

Schwartz J., Marcus A. 1990. Mortality and air pollution in London: A time series analysis. Am J Epidemiol. 131:185-195.

Schwartz, J. 1991. Air pollution and daily mortality in Detroit. Environ Res. 56:204-13.

Schwartz, J., Dockery, D.W. 1992a. Particulate air pollution and daily mortality in Steubenville, Ohio. Am J Epidemiol. 135: 12-19.

Schwartz, J., Dockery, D.W. 1992b. Increased mortality in Philadelphia associated with daily air pollution concentrations. Am Rev Respir Dis. 145:600-4.

Schwartz, J. 1993. Air Pollution and daily mortality in Birmingham, Al. Am J Epidemiol. 137: 1136-47.

Table 4. Risk Ratio of $PM_{10} \geq 50\ \mu g/m^3$ by Cause of Death and Season[4]

| | Respiratory | | Cardiovascular | | Other | | Total | |
|---|---|---|---|---|---|---|---|---|
| Season | N | RR (95% C.I.) | N | RR (95% C.I.) | N | RR (95% C.I.) | N | RR (95% C.I.) |
| Winter | 201 | 1.141(0.8382,1.554) | 904 | 1.215(1.049,1.408) | 771 | 0.930(0.799,1.083) | 1876 | 1.078(0.975,1.191) |
| Spring | 186 | 1.471(0.874,2.477) | 774 | 1.116(0.843,1.477) | 822 | 1.072(0.810,1.418) | 1782 | 1.131(0.940,1.361) |
| Summer | 155 | 0.669(0.344,1.303) | 716 | 1.029(0.795,1.332) | 776 | 0.821(0.615,1.097) | 1647 | 0.904(0.752,1.087) |
| Fall | 167 | 0.826(0.521,1.308) | 817 | 0.899(0.734,1.10) | 836 | 1.063(0.881,1.281) | 1820 | 0.966(0.847,1.102) |
| Total | 709 | 1.034(0.855,1.248) | 3211 | 1.134(1.039,1.238) | 3205 | 0.947(0.864,1.039) | 7125 | 1.039(0.978,1.104) |



[4]From April 7, 1985 - December 31, 1992. Estimated by Poisson regression model adjusted for year, age and temperature.

Table 5. Risk Ratio of $PM_{10} \geq 50\ \mu g/m^3$ by Location of Death and Season[5]

| Season | Hospital N | Hospital RR (95% C.I.) | Nursing Home N | Nursing Home RR (95% C.I.) | Home + Other N | Home + Other RR (95% C.I.) | Total N | Total RR (95% C.I.) |
|---|---|---|---|---|---|---|---|---|
| Winter | 911 | 1.048(0.908,1.209) | 451 | 1.004(0.820,1.228) | 514 | 1.207(0.994,1.465) | 1876 | 1.078(0.975,1.191) |
| Spring | 863 | 1.085(0.832,1.415) | 394 | 0.969(0.627,1.479) | 525 | 1.329(0.964,1.830) | 1782 | 1.131(0.940,1.361) |
| Summer | 752 | 0.811(0.615,1.069) | 372 | 1.044(0.708,1.537) | 523 | 0.955(0.692,1.318) | 1647 | 0.904(0.752,1.087) |
| Fall | 919 | 0.975(0.810,1.174) | 395 | 0.955(0.720,1.267) | 506 | 0.960(0.749,1.230) | 1820 | 0.966(0.847,1.102) |
| Total | 3445 | 1.015(0.930,1.106) | 1612 | 1.028(0.905,1.166) | 2068 | 1.092(0.976,1.220) | 7125 | 1.039(0.978,1.104) |



[5]From April 7, 1985 - December 31, 1992. Estimated by Poisson regression model adjusted for year, age and temperature.

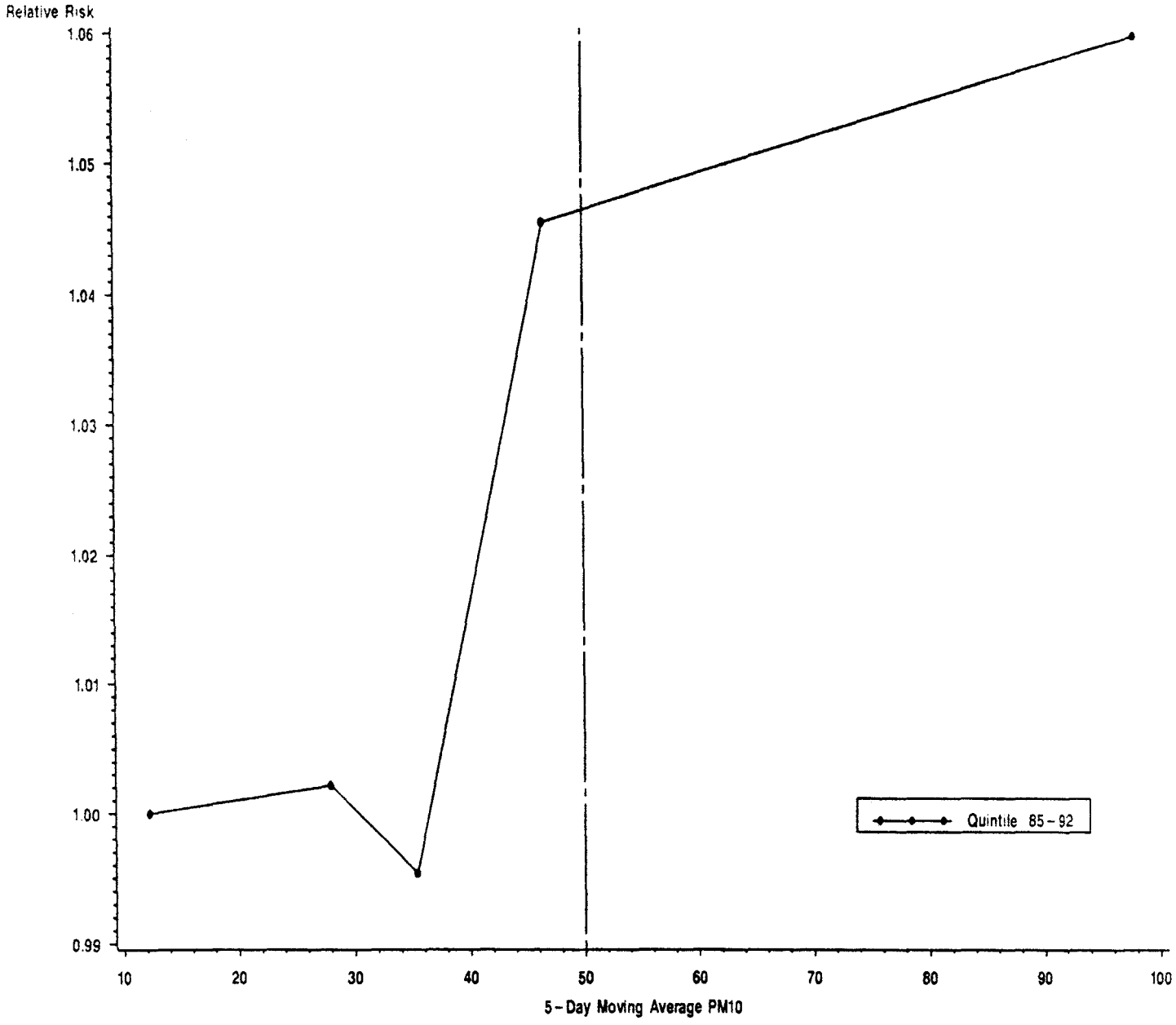


Figure 1 Excess Mortality by Each Quantile of $PM_{10}$ Exposure: Dose-Response Function of $PM_{10}$

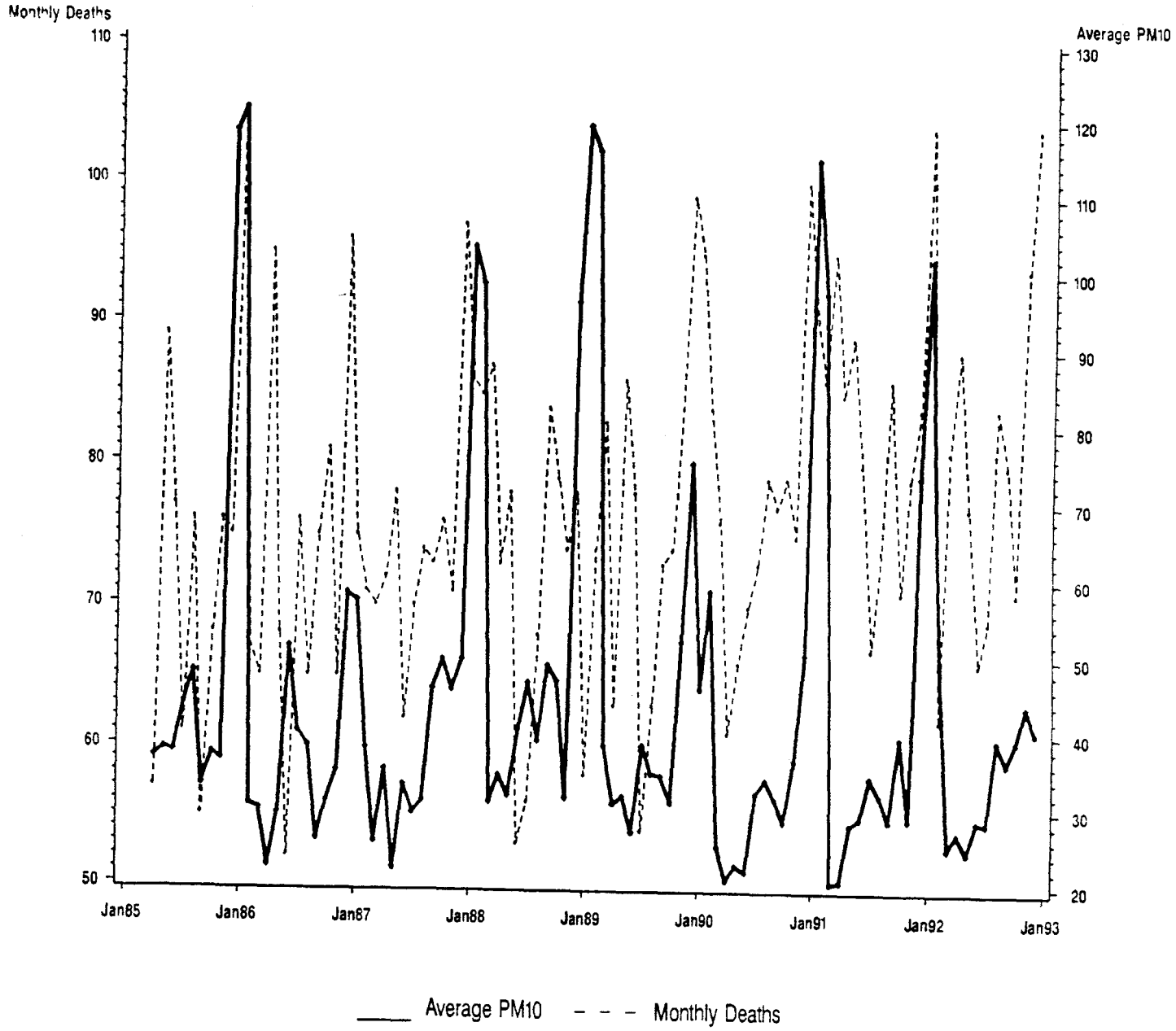


Figure 2 Monthly Deaths and Average $PM_{10}$ Level

Center for Health Statistics for the years 1974 to 1984 inclusive. Deaths due to accidents were excluded. There are two mortality data sets that can be analyzed. The first, which we call 'total' mortality, consists of all deaths (excluding accidents) occurring in the SMSA. The other, which we call 'restricted' mortality, consists of all deaths (again excluding accidents) occurring in the SMSA among individuals who are classified as being residents of the SMSA. It is clear that the restricted mortality data form a subset of the total mortality data. It is not quite clear from their paper which mortality data were used by Schwartz and Dockery in their analysis. We were told by Schwartz (personal communication) that they used restricted mortality. When we compared the distribution of mortality counts with the distribution reported in their paper it was clear that neither mortality data set was identical to the data analyzed by them. However, the distribution of the restricted mortality was closer to the distribution reported in that paper. See figure 1.

Air pollution data were obtained from the Aerometric Information Retrieval Service of the Environmental Protection Agency. Table 1 shows the percentile distribution of the TSP readings in the Steubenville area during the period of observation. It is clear from that table that this distribution is in good agreement with that reported in Schwartz and Dockery. Between 1974 and 1978, sulfur dioxide measurements were made in two distinct ways. A continuous monitoring system, which reported hourly measurements was introduced in 1974. In addition, measurements were also available from an integrated gas bubbler sampler, which was run until 1978. Thus, the sulfur dioxide measurement series over the period covered by the analysis consists of either the continuous or the gas bubbler measurements prior to and including part of 1978, and continuous measurements thereafter. We label the two resulting series of measurements as C-C (to indicate that the continuous monitor was used for the entire period 1974-1984) and GB-C (to indicate that the gas bubbler was used from 1974 to 1978 and the continuous monitor thereafter). Because of problems with the newly introduced continuous monitoring system, Schwartz

and Dockery thought it preferable to use the GB-C series in their analyses. As can be seen from table 1, the percentile distributions of both sulfur dioxide series correspond quite closely to the distribution reported by them, with the C-C series distribution being virtually identical to the distribution in their paper. We used both the GB-C series and the C-C series in our analyses.

Information on daily weather was obtained from the National Oceanic and Atmospheric Administration weather station at the Pittsburgh airport. Summary statistics for the weather variables were quite similar to those reported by Schwartz and Dockery.

Methods of analysis: The data were analyzed using the method of Poisson regression. The regression model included an intercept term and additional terms for the weather and pollution related variables of interest. The parameters of the model were estimated by maximizing the likelihood. Because daily mortality counts from day to day are likely to be correlated, the method of Poisson regression may not yield consistent estimates of variances. A recent method, introduced by Liang and Zeger (1986) can then, in principle, be used for estimating the parameters and obtaining robust standard errors for these estimates. This, so called generalized estimating equation (GEE) approach, introduces a working correlation matrix to describe the correlations among the observations. Liang and Zeger showed that, providing the sample size is large enough, the choice of the working correlation matrix does not affect the consistency of the estimates. We examined the auto-correlation function of the mortality data in Steubenville and found, to our surprise, that there was no evidence that mortality counts on any two days were correlated (see figure 2). Thus, we believe that it is unnecessary to use the GEE approach. In fact, the GEE approach may yield misleading results for this data set, as is discussed below. However, in an effort to confirm published results we used GEE with various working covariance matrices in addition to straight Poisson regression for the analyses of the Steubenville

data. In order to compensate for possible extra-Poisson variation in the data, we computed a 'sandwich' estimator (see expression 7 in Zeger and Liang, (1986)) for the variance when using Poisson regression. We found, however, that this 'robust' estimate of variance was comparable to the 'naive' estimate of the variance obtained from Poisson regression.

Finally, a few words on the use of GEE for the analyses of mortality counts are in order because this method has now been applied to many different data sets. This approach was introduced for the analyses of correlated longitudinal data. Correlations in the data are addressed by way of a working covariance matrix, and if the sample size is large enough (i.e. asymptotically) the method yields consistent estimates of the parameters and the standard errors even if the choice of the working covariance matrix was not particularly accurate to begin with. However, for any given (finite) data set a reasonable choice of the working covariance matrix is always to be preferred to one that is clearly inappropriate. In the context of correlated mortality counts, sample size refers, not to the size of the population, but to the number of independent years of observation. Schwartz and Dockery used an exchangeable covariance structure for their analysis. This choice of working correlation matrix is clearly inappropriate because it assumes, for example, that mortality counts on January 1 and December 31 of a given year are correlated to the same extent as the mortality counts on January 1 and January 2 of that year, and that mortality counts on December 31 of that year and January 1 of the next year are not correlated. In fact, if full year mortality is to be analyzed, then mortality over the entire period of observation should be considered to be a single time series. There is little justification for treating single years as independent units of observation. If seasonal mortality is analyzed, on the other hand, then the assumption of independence would appear to be justified: mortality in the summer of 1979 can be considered to be independent of the mortality of the summer of 1978. But even if analysis is performed by season, there is a sample size problem in the use of GEE with these data. For the asymptotic results to hold, the number

result is in good agreement with the result reported in Schwartz and Dockery (1992a, their table 2).

We found that the effect of TSP is greatly attenuated when sulfur dioxide is included in the model. With full mortality, we find, as shown in table 2 and figure 3 that there is a substantial attenuation of the TSP effect when both TSP and sulfur dioxide are included in the model, the estimated parameter for TSP decreasing from 0.029 (s.e = 0.011) to 0.018 (s.e. = 0.013). If instead of using the full mortality data, we use restricted mortality (used by Schwartz and Dockery in their analysis) and the GB-C sulfur dioxide series, we get essentially the same results. The attenuation of the TSP coefficient is not as large, however (the estimated parameter drops from 0.031, s.e. = 0.012, to 0.025, s.e = 0.014). This is in contrast to the conclusion of Schwartz and Dockery who say, "When both sulfur dioxide and particulates were included in the model simultaneously, particulates remained significant ($\beta = 0.000300 \pm 0.000128$), with little reduction in its estimated coefficient, while sulfur dioxide was insignificant ($\beta = 0.0059 \pm 0.0048$) with a substantial attenuation in the estimated effect." If, however, GEE with an exchangeable correlation structure is used with restricted mortality and the GB-C sulfur dioxide series, both TSP (estimated coefficient = 0.026; s.e. = 0.019) and sulfur dioxide (estimated coefficient = 0.042; s.e. = 0.071) are statistically insignificant. When the C-C sulfur dioxide series is used, we can reproduce the finding of Schwartz and Dockery (TSP is significantly associated with mortality and sulfur dioxide is not), as is shown in figure 4. In summary, we can reproduce Schwartz and Dockery's results only if we use the C-C sulfur dioxide series, a series that they explicitly state they did not use because of measurement error problems.

We also analyzed the mortality data by season. We defined the seasons as follows: Winter: December, January, February; Spring: March, April, May; Summer: June, July,

August; Fall: September, October, November. The results of the analysis using both full and restricted mortality and the GB-C sulfur dioxide series are shown graphically in figure 3. It is quite clear from this figure that the introduction of sulfur dioxide into the model attenuates the coefficient for TSP, and vice versa. We also found a significant effect of temperature in Fall and in Winter, mortality being highest on the coldest days (lowest quintile of temperature). In Summer, mortality is highest on hot days (highest quintile of temperature), however the coefficient is not statistically significant. With the C-C series of sulfur dioxide measurements, the effect of TSP is stronger, whether or not sulfur dioxide is included in the model. Sulfur dioxide, however, appears to have a protective effect in Summer (statistically significant if GEE is used). Thus, the use of the C-C series does lead to some rather unexpected results.

A comparison of figures 3 and 4 shows that the choice of sulfur dioxide series used in the analyses makes a rather large difference to the results. The use of the C-C series leads to results that are much closer to those obtained by Schwartz and Dockery.

DISCUSSION: Regression analyses of the mortality data reveal that by far the most important term in the regression is the intercept. Thus, the weather and pollution variables explain only a very small fraction of the daily deaths. In view of the smallness of the effects we are trying to detect, robustness of results is of great importance. The Steubenville results are not robust, however. This is particularly disturbing in view of the fact that the restricted mortality is a subset of the total mortality data set and differs from it only to the extent of about 0.4 deaths per day on average. Similarly, the GB-C and C-C sulfur dioxide series are identical between mid-1978 and 1984, and as table 1 indicates the distribution of the two series is similar. This fact suggests that the results of the analyses are rather sensitive to measurement errors. The results depend, as well, on whether

Poisson regression or GEE is used for the analysis. For the reasons discussed above, we do not believe that GEE should be used for analyses of these data.

Taken together, the results of the Steubenville analyses are consistent with weak associations between weather and air pollution and mortality. However, in view of the instability of the results, consistency with the results of analyses in other cities must be sought before any firm conclusions can be drawn. Certainly, it would be premature on the basis of these data to single out any single component of air pollution as being responsible. In this report we have considered only two of the six criteria pollutants and have shown that introduction of the second pollutant into a regression model results in an attenuation of the coefficient of the first. There is little *a priori* reason for excluding the other criteria pollutants from the analysis. Future research efforts should be directed at investigating combinations of pollutants.

The role of weather needs to be better investigated as well. Our results suggest that pollutants and season may interact. Kalkstein (1991) has suggested that synoptic categories of weather, which take into account many more weather variables than considered in current analyses, may be more strongly associated with mortality than single weather variables. We are currently investigating these ideas in Steubenville. In particular it would be of great interest to consider synoptic categories for weather together with the criteria pollutants in a single regression model. Such an analysis should ideally be performed in a metropolitan area with a large number of daily deaths, and with many years of observation. If daily mortality counts are correlated, then both the GEE approach and bootstrap methods should be used for the construction of confidence intervals.

Finally, although we attempted to assemble and analyze a data set identical to the data analyzed by Schwartz and Dockery, we were not entirely successful in our efforts.

Although we went to the same primary data sources, we were unable to duplicate exactly either the mortality or the pollution data sets. If the differences in our results are due to the small differences in the data analyzed, they reinforce a point made earlier, namely that the results of these analyses are not robust to small perturbations of the data.

REFERENCES:

**Kalkstein LS**: A new approach to evaluate the impact of climate on human mortality. Environmental Health Perspectives 96: 145-150, 1991.

**Liang KY, Zeger SL**: Longitudinal data analysis using generalized linear models. Biometrika 73: 13-22, 1986.

**Schwartz J**: Particulate air pollution and daily mortality in Detroit. Environmental Research 56: 204-213, 1991.

**Schwartz J**: Air pollution and daily mortality in Birmingham, Alabama. Am J Epidemiol 137: 1136-1147, 1993.

**Schwartz J, Dockery DW**: Particulate air pollution and daily mortality in Steubenville, Ohio. Am J Epidemiol 135: 12-19, 1992a.

**Schwartz J, Dockery DW**: In creased mortality in Philadelphia associated with daily air pollution concentrations. Am Rev Respir Dis 145: 600-604, 1992b.

**Zeger SL, Liang KY**: Longitudinal data analysis for discrete and continuous models. Biometrics 42: 121-130, 1986.

Table 1: Percentile distributions of sulfur dioxide and total suspended particulates (TSP). Refer to the text for definitions of GB-C and C-C. S - D refers to the distributions reported in Schwartz and Dockery (1992).

SULFUR DIOXIDE (ppb) PERCENTILE DISTRIBUTION

| | 10 | 25 | 50 | 75 | 90 | Mean |
|---|---|---|---|---|---|---|
| GB -C | 7 | 13 | 23 | 39 | 58 | 28.8 |
| C - C | 7 | 12 | 23 | 38 | 56 | 28.6 |
| S - D | 7 | 12 | 23 | 38 | 55 | 28 |

TSP (micrograms per cubic meter) PERCENTILE DISTRIBUTION

| | 10 | 25 | 50 | 75 | 90 | Mean |
|---|---|---|---|---|---|---|
| This study | 38 | 58 | 93 | 144 | 212 | 113 |
| S - D | 36 | 56 | 91 | 139 | 209 | 111 |

Table 2: Results of Poisson regression analysis of the Steubenville total mortality data set using the covariates used in a previous analysis (Schwartz and Dockery, 1992a). Indicator variables for winter and spring (defined in text). Hot and humid defined as 24-hour mean temperature > 21.1°C and mean dew point temperature > 18.3 ° C. Coefficient for TSP represents effect of an increase of 100$\mu g/m^3$ in the previous day's mean. Coefficient for $SO_2$ represents effect of an increase of 100ppb in previous day's mean. Individual years were controlled for in the analysis by indicator variables.

| VARIABLE | COEFFICIENT (S.E) | RELATIVE RISK (95% C.I.) |
|---|---|---|
| Intercept | 1.174 (0.032) | |
| TSP | 0.018 (0.013) | 1.018 (0.992 - 1.044) |
| $SO_2$ | 0.072 (0.048) | 1.072 (0.978 - 1.181) |
| Winter | 0.080 (0.022) | 1.080 (1.038 - 1.130) |
| Spring | 0.020 (0.021) | 1.020 (0.979 - 1.063) |
| Hot and humid | 0.023 (0.042) | 1.023 (0.942 - 1.111) |

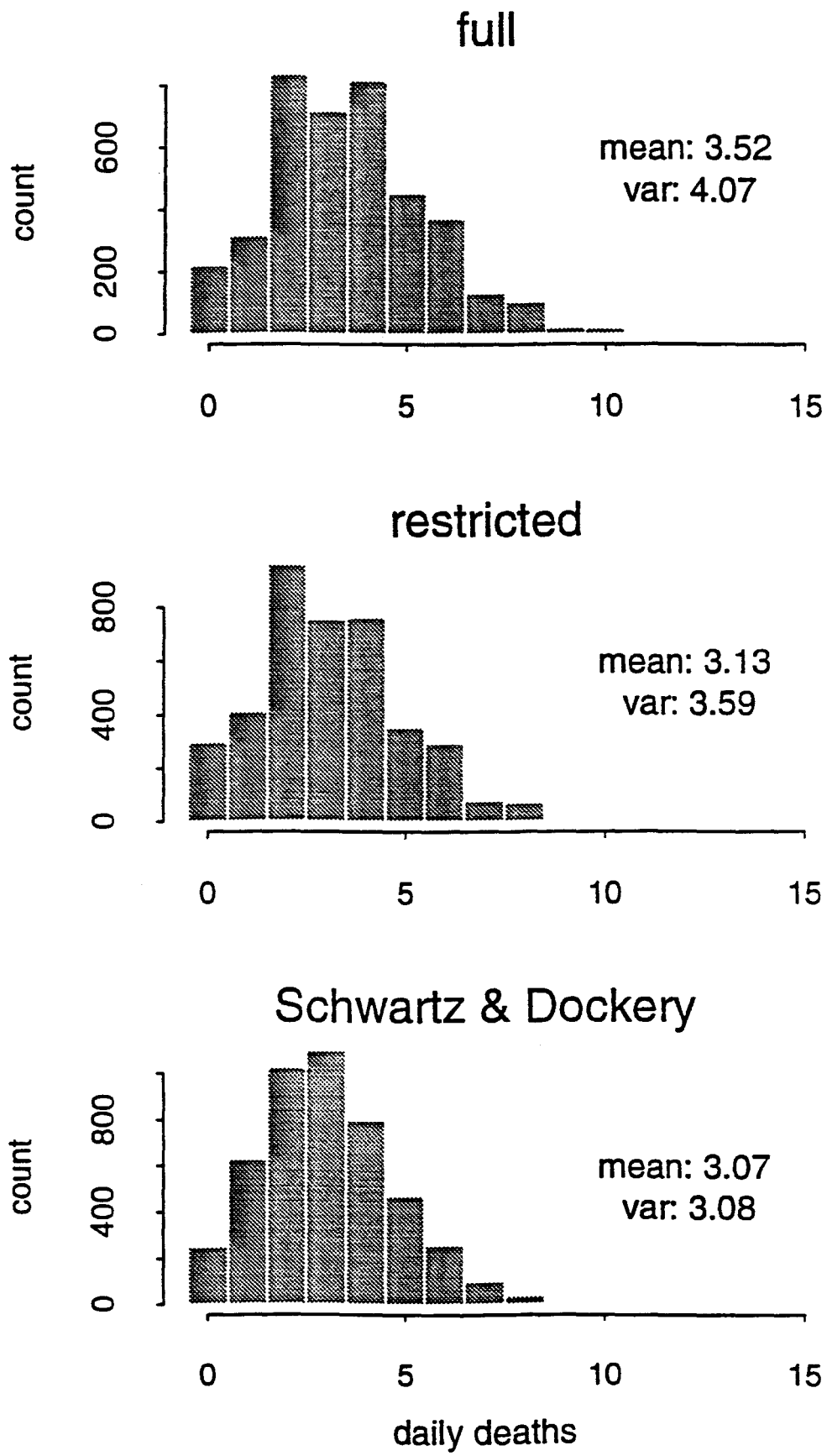


**Figure 1**: Distribution of daily mortality in the Steubenville SMSA between 1974 and 1984. For definitions of 'full' and 'restricted' see text.

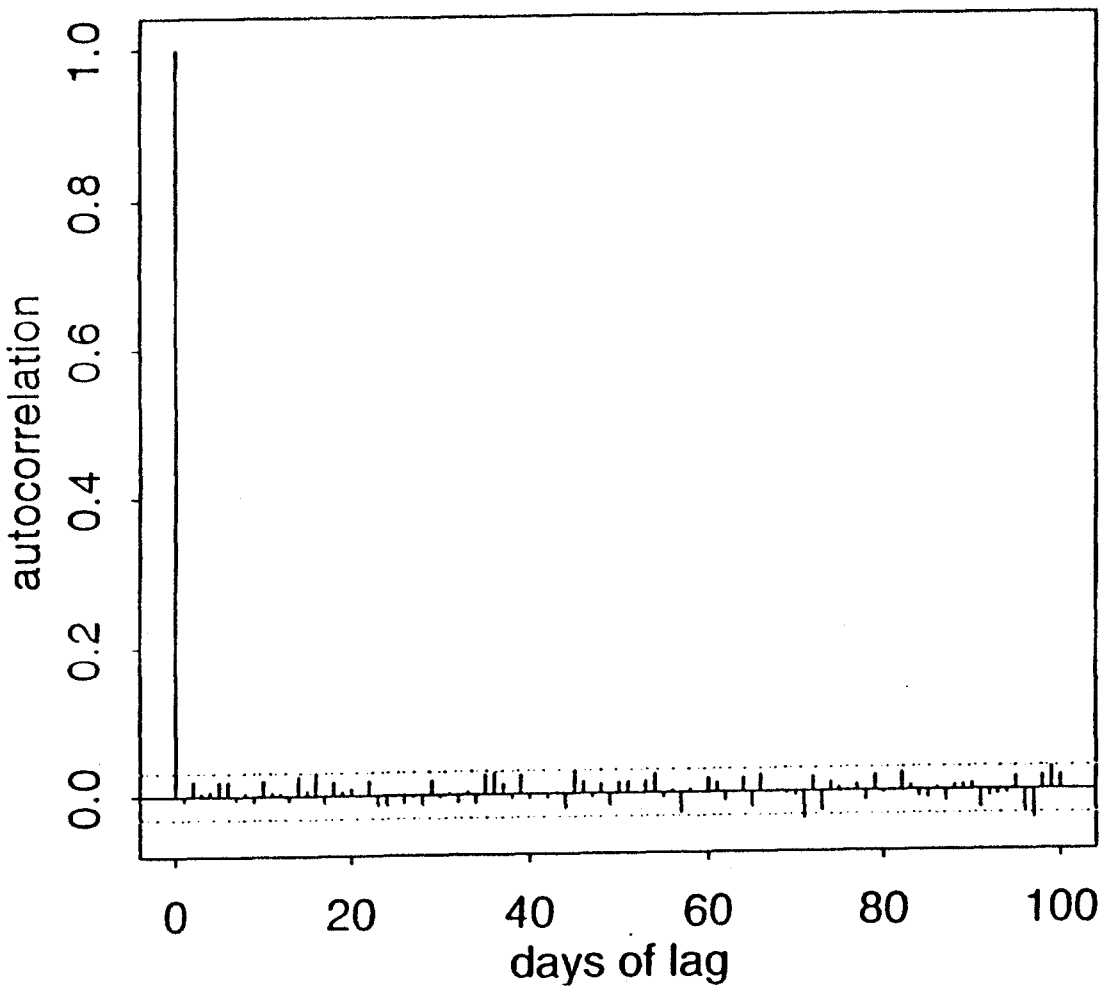


**Figure 2**: Autocorrelation function of the daily total mortality series in Steubenville. Correlation between mortality counts on days separated by various lag times is shown. Thus, correlation for a lag of 0 is 1, indicating that the mortality count on any given day is perfectly correlated with itself The small negative correlation at lag 1 indicates that the mortality counts on any two consecutive days are estimated to have a small negative correlation. The dotted lines represent the 95% confidence intervals for the correlations

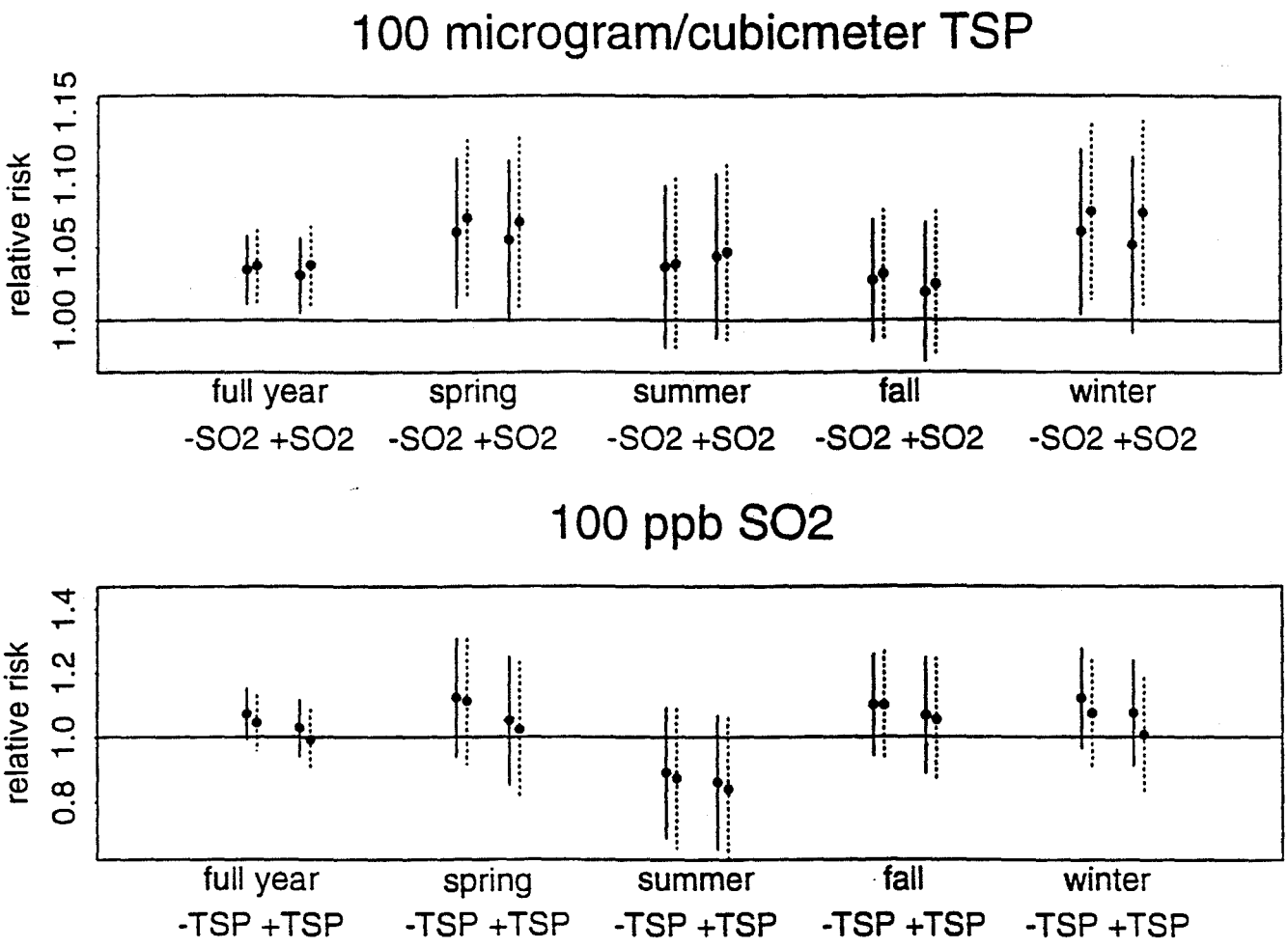


**Figure 4**: Same as figure 3, except that the C-C series of sulfur dioxide measurements is used.

# DAILY MORTALITY ANALYSIS BY USING DIFFERENT REGRESSION MODELS IN PHILADELPHIA COUNTY, 1973-1990

by Yuanzhang Li and H. Daniel Roth

Yuanzhang Li is senior statistician at Roth Associates, Inc.

H. Daniel Roth is president and founder of Roth Associates, Inc.

ABBREVIATED TITLE: PHILADELPHIA MORTALITY

Please address all correspondence to:

Yuanzhang Li, Roth Associates, Inc., 6115 Executive Blvd., Rockville, MD 20852; telephone 301-770-4405; fax 301-770-9248

Philadelphia was analyzed because: (1) the city has a large enough count of daily deaths for a meaningful daily mortality analysis, and (2) daily data on particulate matter and other critical pollutants are available for analysis from sources such as EPA.

Other investigators who have analyzed Philadelphia data are Wyzga (1978) and Schwartz and Dockery (1992a). Our analysis of the Philadelphia data differs from these earlier analyses in several ways. First, we consider a greater number of years of data. Wyzga analyzed the data from 1957 to 1966, and Schwartz and Dockery from 1973 to 1980, as compared to our analysis which considers the data from 1973 to 1990. Also our analysis examines a greater complement of variables. Schwartz and Dockery examined sulfur dioxide and particulate matter; Wyzga looked at only average coefficient of haze (COH). As stated above, we investigated sulfur dioxide, ozone, and particulate matter. Finally, we subjected the data to a larger battery of statistical analyses than were applied by previous investigators in the literature. Wyzga analyzed the data using a regression model; Schwartz and Dockery used the Poisson model.

## DATA

Data on mortality which included the age, the race, and the gender of the descendant and the place of death, the date of death, and the cause of death by ICD code (International Classification of

included: daily $SO_2$ (part of 1973, 1974-1990), daily $O_3$ (part of 1973, 1974-1990), and total suspended particulates (TSP) (1973-1990). The few days for which there were missing data were excluded from the analysis. The distributions of TSP, $SO_2$, $O_3$, maximum temperature, mean humid, dewpoint, and daily mortality by age are summarized in Table 1. Distributions of daily mortality by cause of death are given in Table 2. The correlations among daily mortality for people 65 and older, previous day's air pollutants, and previous day's weather factors are given in Table 3.

Several observations can be made from examining the distributions of the variables in the study (Table 1). To begin with, the levels of particulate matter in Philadelphia probably were in compliance with or close to attainment of the National Ambient Air Quality Standards for particulate matter. The 95 percentile level of TSP levels was 120 $\mu g/m^3$, which is well below the annual National Ambient Air Quality Standard of 260 $\mu g/m^3$ up until 1987. Furthermore, a more detailed examination of the data shows that almost all of the higher levels in the study came from the period of the 1970s. For example, the number of days with TSP levels over 150 $\mu g/m^3$ is 33 from 1973-1974, 29 from 1975-1979, and only 5 from 1980-1990.

The problem of separating out the pollution effects from the weather effects was further complicated by the fact that some weather variables were highly correlated with each other as well as with pollutant variables. For instance, maximum temperature and dewpoint were correlated at the 0.92 level; dewpoint and mean humidity at the 0.54 level; TSP and maximum temperature at 0.26; and ozone with maximum temperature at 0.64.

Finally, the three pollutants considered in the study were also highly correlated with each other. TSP and sulfur dioxide were correlated at the 0.57 level; ozone and sulfur dioxide were correlated at the -0.26 level; and ozone and TSP at 0.18. All of these levels were significant at the .01 level or lower.

## METHODS

Several statistical models exist for analyzing the association between daily mortality and air pollution. Ostro (1984) and Kinney and Ozkaynak (1991) applied a regression model to deviations of moving averages of daily mortality. Hatzakis et al. (1986) used a trigonometric model first to fit the monthly average daily mortality, and used a regression model on the residuals. Shumway et al. (1988) used a regression model on total mortality and a non-linear regression on smoothed mortality. Fairley (1990) used the Poisson-square root regression models on daily mortality. Finally, Schwartz (1991, 1993), Schwartz and Dockery (1992a,b), Dockery et

a logarithmic transformation and a square root transformation. To adjust for additional time trends in the data, we used deviations from moving averages (7, 15, and 29 days), deviations from averages of that day of the year, and deviations from trigonometric curves.

In the model, the current and previous day's pollutant variables and several weather variables (such as temperature, dew-point, and relative humidity) were always considered to be continuous. On the other hand, following the lead of other investigators (e.g. Schwartz and Dockery), we defined the indicated variables such as "cold day," days in which the maximum temperature was in the lower 5% percentile (or less than 31 degrees; see Table 1). "Hot days," "humid days," and "hot-humid days" were defined in a similar manner. Other indicated variables were year, which was incorporated to adjust for long-term changes in daily mortality, and season. In some analyses a continuous variable -- "day of year" -- was used to adjust for time trends within a year. The data were also stratified by season and analyzed separately.

Lastly, following the lead of previous studies, in some analyses the average of two days' air quality was used, and in other analyses just the previous day's levels were considered. Instead of using "barometric pressure" as a variable directly, the increase (or decrease) in barometric pressure from the previous day ($\delta$-barometric) was used because the $\delta$-barometric pressure, which is significant in many models, was found to be more highly correlated

with daily deaths than the individual daily pressure values themselves. For the same reason, we used the difference in precipitation.

## FINDINGS

### The Influence of Season

The significance of the association between pollutants and mortality varied by age, by season, and by the number of pollutants in the model. The results from the Poisson regression with logarithm transformation analysis for the group 65 and older and for the group under 65 are given in Table 5. For the group 65 and older, the t-value of TSP ranged from -0.02 (in winter, when TSP and $SO_2$ are incorporated into the model) to 2.83 (also in winter), (see Table 5). TSP was a statistically significant predictor of mortality only in the winter and only in a model where no other pollutants were considered; the associated coefficient is 0.0005.

Regarding the other seasons, there is no indication of a TSP effect. The t-values in the summer ranged from 0.48 to 1.72; in the spring from 0.60 to 0.98; and in the fall from 0.93 to 0.99. TSP was never a significant predictor of mortality in the under-65 age group. Similarly, there was much variation in the predictive power of other pollutants in the seasonal analyses.

## Analysis by Cause of Death

Total suspended particulates (TSP) was never a statistically significant predictor of mortality in any of our analyses by cause of death (Table 6). In the analysis of individuals 65 and older, the t-values of TSP were always positive for cancer and pneumonia; the t-values of TSP for CVD were consistently negative; and the t-values of TSP for COPD were mixed. The t-value of TSP for COPD was positive in the models where two or more pollutants were considered, but it was negative in the model where TSP alone was analyzed. In the analysis of the under-65 age group, the t-values of TSP for pneumonia and other disease were always negative, and the t-values of TSP for cancer, CVD, and COPD were mixed.

## Averaging Time Considered in the Analysis

Different investigators have used different pollution level averaging assumptions. For example, in Schwartz (1991) daily mortality was compared to previous day's pollution levels; in Schwartz and Dockery (1992a) the mean of pollution in the previous day and the current day was used as the indicator; in Schwartz (1993) the mean of the current and the previous two days was used; in Pope et al. (1992) the mean of the current day and the previous four days was used. In our analysis of the data, we looked at the average from 1 to 4 days (Table 7).

In the analysis, the coefficients and their associated t-values are highest in the model using the mean of two days as the indicator. In the analysis where two days' weather is incorporated, the t-value of TSP is 2.21, 1.31 in the one-day analysis, and 0.38 or 1.02 in the three-day analyses. In addition, the t-values are also much higher in the two-day analysis even when one or two pollutants are considered. We do not know which pollutant averaging time is most desirable, but it is important to consider averaging time when comparing results across studies.

## The Influence of Statistical Models

As we discussed earlier, a host of statistical models were applied to the data (Table 4). Also, three sets of analyses were examined: TSP, $SO_2$, and $O_3$ were incorporated into the model; TSP and $SO_2$ were used; and TSP alone was considered. The resulting t-values from these analyses (in the case where two days' weather variables were incorporated into the model) are given in Table 8 (65 and older) and Table 9 (under 65). Analyses of these tables show that:

1. The significance of a pollutant predictor of daily mortality changes from model to model. For example, in the group 65 and older, the t-value of TSP varies from 0.1 to 6.6.

analysis of eighteen years of Philadelphia data, the results varied by season, the cause of death, the number of pollutants in the model, the model used to analyze the data, and whether or not interactions between pollution variables and weather were considered.

Even in the models with the greatest explanatory power, 18% of variations in daily mortality could be explained, with weather variables and seasonality being the strongest predictors. In analyses using weather alone, about 14% of variations in daily mortality could be explained, but in models using pollution alone, less than 5% could be explained. Also in a two-stage analysis where weather variables are forced into the model and pollution is left to explain the residual, pollution accounts for less than 1% of the variation.

Further complicating the issue is the fact that the pollution variables and the weather variables are all highly correlated with each other, making it difficult to separate out the effect of one pollution variable from another and the effect of weather from pollution. Thus it is difficult to say with any degree of certainty whether those significant associations between TSP and daily mortality observed in our analyses are real or surrogate behavior for some other pollutants or weather variables.

Second, the analysis by season showed no consistent significant association between TSP and mortality. For older people (i.e., 65+), this relationship was significant only in winter in the case where other pollutants were not considered in the equation. In contrast, for younger people (i.e., under 65), the relationship was significant only in the spring in the case where other pollutants were not considered. In both cases (i.e., for older and younger subjects), TSP was insignificant where more than one pollutant was taken into account.

Third, analyses broken down by causes of death indicate that if anything, TSP is negatively correlated with CVD (for the 65 and older group), the disease category with the greatest number of deaths in our analysis. In addition, TSP is negatively correlated with other causes of death, but the associations are never statistically significant. Caution should be followed in overinterpreting the death by cause results because cause of death data are ultimately based on death certificates -- information that can be notoriously misleading.

In summation, a comprehensive analysis of the Philadelphia data does not point to a clear association between daily mortality and particulate matter or to any other pollutant. For almost every result suggesting a positive association between particulate levels

and daily mortality, there are negative or nonsignificant results pointing the other way. It would therefore be incorrect to conclude from a single regression analysis of the Philadelphia data that "X" number of deaths can be assigned to each incremental increase of particulate matter (or any other type of pollutant).

Our analysis also raises some important issues regarding the appropriate methods for analyzing air pollution daily mortality data in the future. First, a determination should be made about the most desirable method for analyzing such data. Following the lead of others, we have used Poisson models, regression models, and autoregressive models, and arrived at results ranging from positive to negative associations. Each of the models that we used has its strengths and weaknesses. In the future, more work should be done to develop optimum models for analyzing mortality data and new approaches should be explored.

Second, a determination should be made about the number of days of pollution and weather measurements to be taken into account in the analysis. As was shown in Table 8, the significance of the predictive power of TSP on mortality is highly dependent on the number of pollution days in the model.

Our analyses clearly indicate that results are data- and model-dependent. Moreover, it is unclear which statistical model generates the most definitive results. Results vary by the model

choice, pollution variable choice, season, and cause. A simple conclusion on the association between daily mortality and air pollution based on a specified model might be biased.

ACKNOWLEDGEMENT

The authors thank the Electric Power Research Institute for their support of this study.

Ozkaynak, Haluk; Thurston, George D. *Associations between 1980 U.S. mortality rates and alternative measures of airborne particle concentration.* Risk Analysis 7(4):449-460, 1987.

Pope, C.A.; Schwartz, J.; Ransom, M. *Daily mortality and $PM_{10}$ pollution in Utah Valley.* Arch. Environ. Health 42:211-217, 1992.

Schwartz, J. *Air pollution and daily mortality in Birmingham, Alabama.* Am. J. Epid. 137(10), 1993.

Schwartz, J. *Particulate air pollution and daily mortality in Detroit.* Environ. Res. 56:204-213, 1991.

Schwartz, J.; Dockery, D.W. *Increased mortality in Philadelphia associated with daily air pollution concentrations.* Am. Rev. Respir. Dis. 145:600-604, 1992a.

Schwartz, J.; Dockery, D.W. *Particulate air pollution and daily mortality in Steubenville, Ohio.* Am. J. Epidemiol. 135:12-19, 1992b.

Schwartz, J.; Spix, C.; Wichmann, H.E., et al. *Air pollution and acute respiratory illness in five German communities.* Environ. Res. 56:1-14, 1991.

Shumway, R.H.; Azari, A.S.; Pawitan, Y. *Modeling mortality fluctuations in Los Angeles as functions of pollution and weather effects*. Env. Res. 45:224-241, 1988.

Wyzga, R.E. *The effect of air pollution upon mortality: a consideration of distributed lag models*. JASA 73(363):463-472, September 1978.

TABLE 1

DISTRIBUTION OF DAILY MORTALITY, POLLUTANTS*, AND WEATHER FACTORS (1973-1990)

| | 5% | 10% | 25% | 50% | 75% | 90% | 95% | MEAN | STANDARD DEVIATION |
|---|---|---|---|---|---|---|---|---|---|
| TSP ($\mu g$) | 32 | 37 | 48 | 64 | 83 | 105 | 120 | 68.51 | 27.91 |
| $SO_2$ (ppb) | 4.10 | 5.89 | 8.93 | 13.91 | 21.80 | 31.83 | 39.46 | 16.98 | 11.87 |
| $O_3$ (ppb) | 1.82 | 3.54 | 8.33 | 17.08 | 28.43 | 39.36 | 46.83 | 19.77 | 14.35 |
| MXTP °F | 31 | 37 | 48 | 63 | 78 | 84 | 86 | 64.06 | 18.88 |
| MNHUMID | 43.75 | 48.88 | 56.88 | 66.00 | 74.17 | 80.38 | 83.29 | 67.01 | 13.41 |
| DEWP °F | 10.00 | 16.75 | 28.50 | 44.25 | 59.17 | 66.42 | 70.17 | 42.90 | 19.04 |
| DAILY DEATH <65 | 12 | 13 | 16 | 19 | 22 | 26 | 28 | 19.06 | 5.03 |
| DAILY DEATH ≥65 | 24 | 26 | 30 | 35 | 40 | 45 | 48 | 35.15 | 7.13 |

* The air quality data is the average of values across Philadelphia.

TABLE 2

DISTRIBUTION OF CAUSE-SPECIFIED DAILY MORTALITY COUNTS
IN PHILADELPHIA COUNTY, 1973-1990

| CAUSE | MORTALITY, DEATHS/DAY* | 5% | 10% | 25% | 50% | 75% | 90% | 95% | Mean | Std |
|---|---|---|---|---|---|---|---|---|---|---|
| Cancer | Age <65 years | 2 | 3 | 4 | 5 | 7 | 9 | 10 | 5.66 | 2.42 |
| | Age ≥65 years | 3 | 4 | 6 | 5 | 10 | 12 | 13 | 7.76 | 2.91 |
| CVD | Age <65 years | 2 | 2 | 3 | 5 | 7 | 8 | 10 | 5.22 | 2.43 |
| | Age ≥65 years | 10 | 12 | 14 | 17 | 21 | 24 | 26 | 17.58 | 4.76 |
| Pneumonia | Age <65 years | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 0.38 | 0.64 |
| | Age ≥65 years | 0 | 0 | 0 | 1 | 2 | 3 | 4 | 1.31 | 1.23 |
| COPD | Age <65 years | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0.23 | 0.48 |
| | Age ≥65 years | 0 | 0 | 0 | 0 | 1 | 2 | 3 | 0.75 | 0.96 |
| Other | Age <65 years | 3 | 4 | 5 | 7 | 9 | 12 | 13 | 7.58 | 3.03 |
| | Age ≥65 years | 3 | 4 | 6 | 8 | 10 | 12 | 13 | 7.75 | 3.04 |

TABLE 3

PEARSON CORRELATION COEFFICIENTS OF THE PREVIOUS DAY'S AIR FACTORS AND DAILY MORTALITY (AGE ≥65) (1973-1990)

| | $SO_2$ | $O_3$ | TSP | MXTP | DEWP | MNHUMID |
|---|---|---|---|---|---|---|
| DEATH* | 0.179+ | -0.143+ | 0.051+ | -0.283+ | -0.263+ | -0.005 |
| $SO_2$ | | -0.260+ | 0.566+ | -0.248+ | -0.264+ | -0.035+ |
| $O_3$ | | | 0.179+ | 0.635+ | 0.488+ | -0.161+ |
| TSP | | | | 0.262+ | 0.211+ | 0.046+ |
| MXTP | | | | | 0.921+ | 0.201+ |
| DEWP | | | | | | 0.535+ |

* Death: daily mortality for persons age ≥65.

\+ Statistically significant (p-value = <0.01)

TABLE 5

THE t-VALUES* IN THE SEASONAL ANALYSIS
MODEL: POISSON-LOG
FACTORS: YEAR, DAY OF SEASON, TWO DAYS' WEATHER

| | AGE | | | | | | |
|---|---|---|---|---|---|---|---|
| | >65 | | | | <65 | | |
| SEASON | TSP | $SO_2$ | $O_3$ | ** | TSP | $SO_2$ | $O_3$ |
| Fall | 0.99 | 0.49 | 1.29 | | -1.16 | 2.75 | -0.15 |
| Winter | 0.05 | 2.55 | 0.92 | | 0.65 | 0.45 | 0.00 |
| Spring | 0.60 | 0.84 | 2.73 | | 0.76 | -0.26 | -0.38 |
| Summer | 0.48 | -0.53 | 2.67 | | 0.05 | 2.56 | -0.34 |
| Fall | 0.93 | 0.45 | | | -0.76 | 2.09 | |
| Winter | -0.02 | 2.68 | | | 0.01 | 1.02 | |
| Spring | 0.51 | 0.58 | | | 0.77 | -0.23 | |
| Summer | 1.59 | -0.94 | | | 0.79 | 1.92 | |
| Fall | 1.19 | | | | 0.68 | | |
| Winter | 2.83 | | | | 0.96 | | |
| Spring | 0.98 | | | | 2.04 | | |
| Summer | 1.72 | | | | 1.89 | | |

* All t-values in this analysis could be considered subject to standard normal distribution because the degree of freedom is higher than thousands.

** Blank means the pollution is not considered in the model.

TABLE 6

THE t-VALUES OF POLLUTANTS IN THE ANALYSIS BY CAUSE
MODEL: POISSON-SQUARE ROOT
FACTORS: YEAR, DAY OF SEASON, TWO DAYS' WEATHER

| | AGE | | | | | | |
|---|---|---|---|---|---|---|---|
| | ≥65 | | | | <65 | | |
| CAUSE | TSP | $SO_2$ | $O_3$ | ** | TSP | $SO_2$ | $O_3$ |
| Cancer | 0.67 | -0.73 | 0.20 | | -0.88 | 1.36 | 1.79 |
| CVD | -1.74 | 1.73 | 3.97 | | -0.46 | 1.37 | 2.32 |
| Pneumonia | 0.88 | -0.40 | 0.65 | | -1.19 | 1.44 | -0.03 |
| COPD | 0.08 | -0.64 | 0.06 | | 0.14 | -0.69 | -1.62 |
| Other | -0.07 | 1.82 | 1.45 | | -0.56 | -0.18 | 1.21 |
| Cancer | 0.93 | -1.06 | | | -0.48 | 0.76 | |
| CVD | -0.74 | 0.56 | | | 0.12 | 0.65 | |
| Pneumonia | 0.61 | 0.53 | | | -1.18 | 1.41 | |
| COPD | 0.23 | -0.66 | | | -0.07 | -0.63 | |
| Other | -0.04 | 1.53 | | | -0.60 | 0.05 | |
| Cancer | 0.39 | | | | 0.14 | | |
| CVD | -0.24 | | | | 1.36 | | |
| Pneumonia | 1.12 | | | | -0.82 | | |
| COPD | -0.37 | | | | -0.65 | | |
| Other | 0.89 | | | | -0.36 | | |

* All t-values in this analysis could be considered subject to standard normal distribution because the degree of freedom is higher than thousands.

** Blank means the pollution is not considered in the model.

TABLE 7

t-VALUES OF POLLUTANT INDICATORS GENERATED
BY DIFFERENT AVERAGING TIME ANALYSES
MODEL: POISSON-LOG
AGE ≥65

| MODEL | SIMULTANEOUS ANALYSIS | | | TSP ONLY |
|---|---|---|---|---|
| | TSP | $SO_2$ | $O_3$ | |
| MP1(2)* | 1.31 | 3.49 | 3.81 | 4.24 |
| MP2(2) | 2.21 | 3.59 | 4.34 | 5.97 |
| MP3(2) | 0.38 | 1.98 | 3.07 | 3.0 |
| MP4(2) | 0.20 | 1.56 | 2.35 | 2.44 |
| MP3(3) | 1.02 | 1.72 | 3.96 | 3.45 |
| MP4(4) | 1.19 | 1.38 | 3.82 | 3.46 |

* MPx(y) stands for mean pollution of current and prior (x-1) days with y days' weather. For example, MP1(2) is prior day's pollution with 2 days' weather.

TABLE 8

t-VALUES OF POLLUTANTS IN DIFFERENT STATISTICAL MODELS: AGE ≥65

| | POLLUTANTS IN THE MODEL | | | | | |
|---|---|---|---|---|---|---|
| Model | TSP | $SO_2$ | $O_3$ | TSP | $SO_2$ | TSP |
| PL | 1.2 | 3.4 | 3.8 | 1.9 | 2.4 | 4.0 |
| PSR | 1.3 | 3.5 | 3.7 | 2.0 | 2.6 | 4.3 |
| AR(6) | 2.4 | 3.8 | 5.8 | 3.6 | 2.7 | 6.6 |
| RFT | 0.8 | 3.0 | 3.2 | 1.5 | 1.9 | 3.1 |
| RDY | 1.4 | 1.6 | 3.1 | 2.0 | 0.8 | 3.1 |
| RDM(7)* | 0.1 | 1.6 | 1.7 | 0.3 | 1.0 | 1.0 |
| RDM(15) | 0.6 | 2.2 | 2.8 | 0.9 | 1.6 | 2.4 |
| RDM(29) | 1.1 | 2.1 | 3.3 | 1.7 | 1.1 | 3.2 |

* 7, 15, 29 = 7, 15, or 29-day moving average.

TABLE 9

t-VALUES OF POLLUTANTS IN DIFFERENT STATISTICAL MODELS: AGE <65

| | POLLUTANTS IN THE MODEL | | | | | |
|---|---|---|---|---|---|---|
| Model | TSP | $SO_2$ | $O_3$ | TSP | $SO_2$ | TSP |
| PL | -0.04 | 2.33 | 1.67 | 0.09 | 1.96 | 1.92 |
| PSR | 0.04 | 2.75 | 1.73 | 0.23 | 2.30 | 2.43 |
| AR6 | 1.11 | 3.00 | -1.06 | -1.90 | -0.94 | -0.66 |
| RFT | -0.77 | 2.45 | 1.52 | -0.50 | 1.89 | 1.33 |
| RDY | -0.15 | 1.38 | 1.40 | 0.02 | 0.96 | 1.32 |
| RDM(7)* | -1.34 | 2.10 | 1.65 | -1.55 | 1.97 | -0.06 |
| RDM(15) | -0.55 | 1.69 | 1.97 | -0.61 | 1.50 | 0.73 |
| RDM(29) | 1.09 | 2.11 | -1.73 | -4.89 | 4.46 | -2.25 |

* 7, 15, 29 = 7, 15, or 29-day moving average.

# A Sensitivity Analysis of Mortality/$PM_{10}$ Associations in Los Angeles

Patrick L. Kinney, Kazuhiko Ito, and George D. Thurston

New York University Medical Center
Nelson Institute of Environmental Medicine
Long Meadow Road
Tuxedo, NY

Running title: Mortality/$PM_{10}$ Sensitivity

Corresponding Author: Patrick L. Kinney, Sc.D.
Nelson Institute of Environmental Medicine
Long Meadow Road
Tuxedo, NY 10987

## INTRODUCTION

There have been several studies published since 1990 that have reported statistically significant and quantitatively similar associations between daily numbers of deaths and gravimetric particulate matter concentrations, either $PM_{10}$ (mass concentrations of particles with aerodynamic diameters less than 10 μm) or TSP (total suspended particulate mass concentration), in metropolitan areas (Schwartz, 1991; Schwartz and Dockery, 1992a, 1992b; Dockery, Schwartz, and Spengler, 1992; Pope, Schwartz, and Ransom, 1992; Schwartz, 1993). It is remarkable that consistent results have been obtained in diverse cities that vary in population, weather patterns, and levels of co-pollutants. The quantitative consistency of the reported associations, and their coherence with epidemiologic studies of morbidity outcomes, have led Schwartz to conclude that the effects of low-level airborne particle exposures on mortality are likely to be causal (Schwartz, 1993).

This body of results has been viewed by many with skepticism. Primary among the concerns raised has been the lack of biological plausibility. It is difficult to understand how a 10 $\mu g/m^3$ increase in $PM_{10}$ levels could result in about a 1% increase in daily deaths in a city, as would be implied from the data. Further, there are no animal toxicology data for particulate matter that suggest significant biological mechanisms that could account for health effects at the low levels of exposure encountered in the population-based studies. The other principal concern has been that the results might be due to an artifact of the complex and specialized statistical methods utilized in most of the recent literature on this topic.

Time series studies of daily mortality and air pollution use a variety of complex statistical methods designed to address the special character of serial data. Because many of the statistical methods employed are relatively new and specialized, they are fully understood by few analysts and even fewer users of time-series results. In addition, there is currently no firm consensus among analysts as to which statistical

A final issue relates to the distribution of the residual variability in mortality that remains after the model is fit. Because death is a discrete event, daily mortality counts will tend to have a Poisson distribution rather than the normal Gaussian distribution that is assumed in most standard analyses. It has been argued that methods which assume Poisson residuals are therefore necessary (Schwartz, 1993). On the other hand, when the mean is large, a Poisson process will be nearly Gaussian. It is not yet clear how much difference this issue makes in practice.

This paper evaluates the sensitivity of mortality/$PM_{10}$ relationships over a range of statistical approaches to the four issues introduced above. Rather than attempting to dictate the single, most valid set of methods, we pose the question, "how much difference does it make to the results when the full range of commonly used methods are tested"? The sensitivity analysis was performed using data from Los Angeles county for the period 1985-1990, the period during which $PM_{10}$ data first became available.

## METHODS

Daily counts of total deaths which occurred in Los Angeles County in the period January 1, 1985 to December 31, 1990 were obtained from National Center for Health Statistics death certificate tapes. Deaths due to accidents and suicides and non-resident deaths were excluded from the total counts, yielding the daily death count variable used in the analyses reported here. $PM_{10}$, $O_3$, and CO data collected in Los Angeles county were obtained in digital form from the US. Environmental Protection Agency's Aerometric Information and Retrieval System (AIRS). 24-hr average $PM_{10}$ concentrations, collected every 6 days, were taken from 4 monitoring sites. Daily maximum 1-hr $O_3$ and CO levels were obtained from 8 sites each. The multiple site data for each pollutant were averaged after filling missing values using a multiple regression algorithm. Missing values that were filled in this way represented 7%, 3%,

sine and cosine variables gave somewhat larger $PM_{10}$ relative risks. There was some indication from these latter results that the mortality/$PM_{10}$ association was stronger in summer than in winter. As a whole, however, the $PM_{10}$ relative risks exhibited a substantial consistency across a range of alternative methods aimed at controlling the cyclic behavior of the data.

Among the cyclic control methods we explored, three yielded DW statistics above 1.8: the model that included 10 sine and cosine variables, and the two models restricted to the summer season. On the basis of these results, we chose to employ the 10 sine/cosine model as the basic approach for cyclic control in all the models discussed below.

Regressions with varying levels of control for temperature and relative humidity (none; same-day; extensive lagged variables) all yielded similar results (Table 1 and Figure 2), although the $PM_{10}$ relative risk, and statistical significance, was reduced somewhat with the more extensive weather controls.

Regression models that included alternative pollutants ($PM_{10}$, lag 1 $O_3$, or CO) all yielded significant (or nearly so) relative risks for the individual pollutants (Table 1 and Figure 3). In bivariate regressions (i.e., regressions including $PM_{10}$ and one other pollutant), results were more variable. With both $PM_{10}$ and $O_3$ in the model, the $PM_{10}$ relative risk was essentially unchanged (RR=1.05), while the $O_3$ relative risk dropped to one. This suggests that the $O_3$ effect on mortality, if any, is weaker than that of $PM_{10}$. The correlation of the slope estimates for $O_3$ and $PM_{10}$ was -0.5, indicating a substantial collinearity in these two pollutants when included simultaneously in the model. In contrast to the situation for $O_3$, the relative risks for <u>both</u> $PM_{10}$ and CO dropped somewhat when both were included in the model, suggesting a similar strength of association with mortality for the two pollutants. There was a moderate collinearity in the slope estimates for these two pollutants (slope correlation = -0.4). Overall, the range

of $PM_{10}$ relative risks observed in these alternative models (1.03 - 1.05) again demonstrated a rather mild degree of sensitivity.

Finally, we evaluated the sensitivity of mortality/$PM_{10}$ associations under differing assumptions about functional form and residual distributions (Table 1 and Figure 4). Three approaches were used: ordinary least squares (assumes a linear relationship and normally distributed errors with constant variance), log-linear regression (assumes an upward curving relationship between mortality and $PM_{10}$ and normally distributed errors with variance that increases in proportion to level of mortality), and Poisson regression (same assumptions as log-linear except that errors are assumed to be Poisson distributed). No difference was observed in the $PM_{10}$ relative risks from the three models.

## DISCUSSION AND CONCLUSIONS

We have evaluated the sensitivity of daily mortality/$PM_{10}$ associations to a range of analytical methods in a newly developed 6-year data set from Los Angeles county. We found that the estimated proportional increase in daily mortality associated with a 100 $\mu g/m^3$ increase in $PM_{10}$ concentration (or relative risk) fell generally between 1.03 and 1.05, regardless of the method used. These results indicate that sensitivity to methods was low in this data set.

These results for Los Angeles represent the first independent confirmation of the mortality/particulate matter associations reported for $PM_{10}$ and TSP in the recent series of papers by Schwartz and colleagues. The $PM_{10}$ relative risk (RR) we obtained for Los Angeles (1.05) is somewhat smaller than those reported previously in Utah Valley (RR=1.16) by Pope and colleagues (1992), St. Louis, MO (RR=1.16) by Dockery and colleagues (1992), and Birmingham, AL (RR=1.11) by Schwartz (1993). Two of those studies used multi-day averages of $PM_{10}$ as the exposure metric, which may have yielded a larger effect by picking up both same-day and lagged $PM_{10}$ effects. We were

Table 1. Descriptive Statistics, Overall and by Season

| Period | Statistic | Deaths (per day) | $PM_{10}$ ($\mu g/m^3$) | $O_3$ (ppb) | CO (ppm) | T (Deg. F) | RH (%) |
|---|---|---|---|---|---|---|---|
| All | Mean | 153 | 58 | 70 | 4.7 | 70 | 70 |
| | Std. Dev. | 20 | 23 | 41 | 2.9 | 7 | 16 |
| | Range | 113-224 | 15-177 | 3-201 | 1-13 | 54-98 | 14-97 |
| Nov-Feb | Mean | 169 | 61 | 36 | 7.4 | 67 | 63 |
| | Std. Dev. | 20 | 29 | 20 | 2.7 | 7 | 20 |
| | Range | 132-224 | 15-177 | 3-100 | 2-13 | 54-88 | 14-97 |
| Jun-Sep | Mean | 143 | 60 | 101 | 2.8 | 73 | 77 |
| | Std. Dev. | 12 | 18 | 37 | 1.5 | 4 | 8 |
| | Range- | 116-179 | 20-116 | 21-201 | 1-8 | 63-87 | 49-90 |

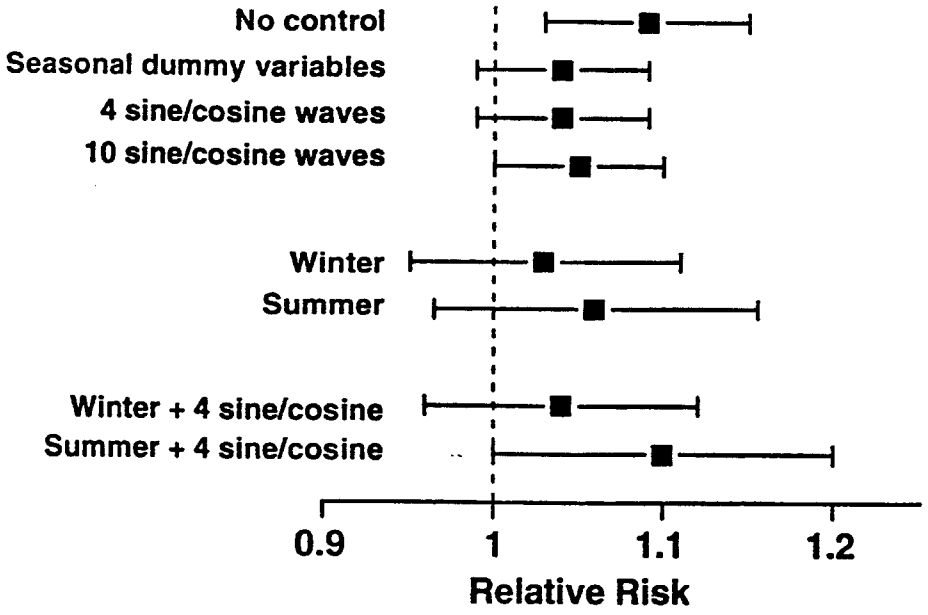


Figure 1. $PM_{10}$ relative risks and 95% confidence intervals obtained with alternative methods to control for temporal cycles. All models included same-day temperature and relative humidity as covariates. Relative risks were computed for a 100 $\mu g/m^3$ increase in $PM_{10}$ concentration.

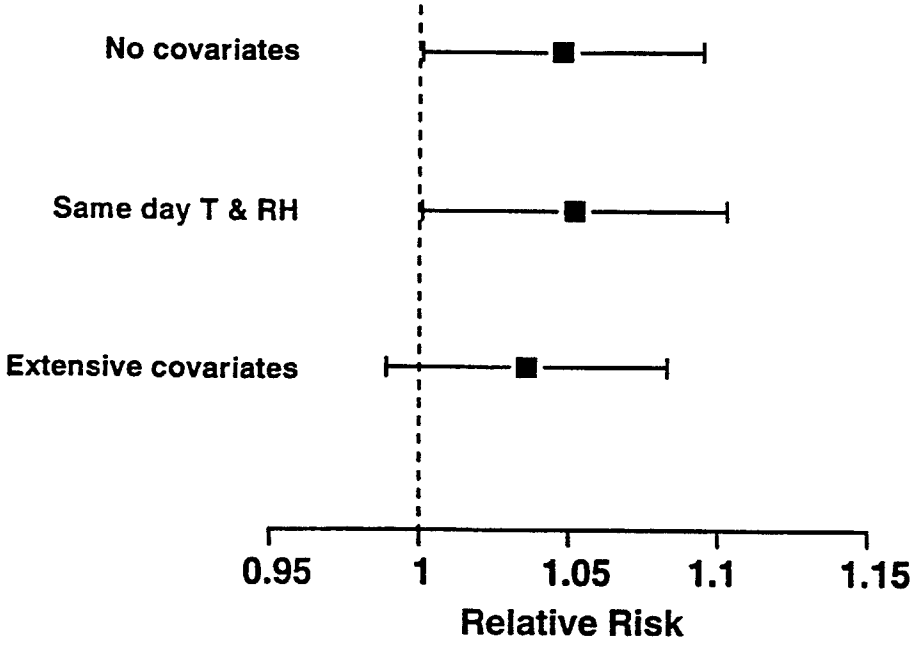


Figure 2. $PM_{10}$ relative risks and 95% confidence intervals obtained with varying levels of weather controls. All models included 10 sine/cosine functions as covariates. Relative risks were computed for a 100 $\mu g/m^3$ increase in $PM_{10}$ concentration.

for submission to Inhalation Toxicology

# Uncertainties in Identifying "Responsible" Pollutants in Observational Epidemiology Studies

Frederick W. Lipfert*
Environmental Consultant
23 Carll Court
Northport, NY 11768

Ronald E. Wyzga
Electric Power Research Institute
P.O. Box 10412
Palo, Alto, CA 94303

*person to whom correspondence should be addressed

Abstract

Studies of community air pollution must deal with a complex mixture of substances, for which the available data on concentrations and their distributions vary greatly in completeness and accuracy. The monitoring database available for some pollutants (such as suspended particulate matter) far exceeds that available for others (such as carbon monoxide or nitrogen dioxide) in terms of spatial and temporal coverage. Little or no routine monitoring data are available on aeroallergens or on particles classified by size and chemistry, for example. In addition, the relationships between outdoor air concentrations and personal exposures vary by chemical species. This paper addresses the concern that the availability and quality of observed data may limit the validity of the conclusions that can be derived from retrospective studies.

The basic assumptions of multiple regression analysis, the statistical tool most commonly used to study the effects of air quality on health, are reviewed. We show by data simulation and by numerical experiments with mortality and air quality data from Philadelphia that differences in the reliability of exposure estimates can be critical in the implied relationships between correlated variables in multiple (joint) regressions. Further, measurement error obscures the true degree of collinearity that may actually be present. Finally, we consider how nonlinear transformations can affect judgments about the relative importance of the variables considered. While models based on linear pollution relationships may be facile and may be convenient in characterizing effects, we have no assurance that they are in fact correct. Resolution of these issues will require better population-based air quality monitoring data as well as laboratory studies appropriate to characterizing the nature of the implied biological responses to the mixtures and concentrations that currently comprise community air quality.

---

*This research was sponsored by the Electric Power Research Institute.*

## INTRODUCTION

Background. The observational or "ecological" epidemiological study is one study design that can be used to study irreversible endpoints such as mortality or hospitalization. At current air quality levels in the United States, even the severest air pollution health effects are considered "weak" from a statistical perspective (Wynder, 1987), because they generally involve individual risk ratios less than 1.5. Thus, strict adherence to statistical requirements is required in order to derive even valid qualitative conclusions. However, because large populations may be exposed to the risks associated with community air pollution, the findings of such observational studies may be important from the perspectives of public health and environmental policy.

Observational studies are intended to deduce relationships between population health responses and environmental characteristics. There are often two inter-related objectives: to identify associations between specific diseases and specific agents, and to define the concomitant dose-response functions and/or "safe" concentration levels. "Safe" levels may be used to establish ambient air quality standards, and dose-response functions are needed to estimate the economic or health benefits that might accrue from new policies to control emissions. In general, there is no longer interest in a study outcome that only shows that air pollution may be "harmful to health." Such harm has been unequivocably established by the air pollution episodes of past decades (Lipfert, 1994a) and by the many laboratory experiments that have followed. Current interest centers on the details of relationships between air pollution and biological responses; this requires separating the effects of collinear pollutants and determining the functional forms of dose-response functions, including whether no-effect thresholds are present. At the risk of pedantry, we first review some of the fundamental concepts involved in deducing

such relationships from experimental data. Much of the following material and notation follows that of Snedecor and Cochran (1967).

## METHODOLOGICAL ISSUES

Statistical Methods. Statisticians have devised many ways to study relationships embodied in a data set. All of them involve comparing a calculated statistic with a range of values that would be expected to occur just to chance (random) variation. If the statistic falls outside of this range, it is said to be "significant" and we then look for explanations for this behavior. The limits of chance variation are established from the distributional properties associated with the statistic, such as those of the Poisson or Gaussian (normal) distributions. It follows that certain assumptions must be met for such comparisons to be valid. Sometimes one or more variables must be mathematically transformed in order to meet these requirements.

For large cities, the data describing community health responses to air pollution may be treated as quasi-continuous, rather than as categorical. The most common measures of association are correlation and regression. Tests of significance with respect to the **existence** of a relationship, using either (bivariate) regression or correlation, require only that the dependent variable be normally distributed. This test compares either the regression slope (b) or the correlation coefficient (r) with zero. When r is not zero and we wish to establish its confidence limits (for the purpose of model comparison, for example), both variables must be normally distributed.

However, the bivariate situation is rare in studies of practical interest, in part because the effects of interest are weak, and thus there are always intervening variables that could confound the outcome if not controlled. As a result, multiple regression is usually required, with

epidemics. For comparisons between communities, we must add factors that describe lifestyle and basic health status differences, including smoking habits, education, diet, exercise, occupational exposure, selective migration, among others. However, there are also certain non-physiological factors that have been shown to contribute, even though biological explanations may be lacking; some of these are variations according to the day of the week, changes in meteorology, and even proximity to the decedent's birthday (Blakeslee, 1992) and cultural beliefs (Blakeslee, 1993). Differences in the "pace of life" have been implicated in differences in health status among communities (Levine, 1990).

In addition, if the effects of air pollution are to be considered on a physiological (rather than psychological) basis, we must estimate actual exposures to pollution. This exposure will vary with the daily activity patterns of the subjects and the nature of all the micro-environments to which they are exposed each day. The only data we have on air pollution across an entire community are the measurements routinely collected from fixed-base monitoring stations, which will have varying relationships to these microenvironments depending on the pollutant, time of day, and the age and health status of the subject, among other factors. The relationships between air monitoring data and the actual exposures of subjects constitute potential sources of bias and of errors in the independent variables. The effects of such errors will be addressed subsequently.

Because we cannot account for all factors that may deterministically affect health outcomes, our model is not in fact described by Eq. [1], but by

$$Y = b_0 + b_1x_1 + b_2x_2 \ldots\ldots + b_nx_n \quad + b_kx_k + e' \qquad [2]$$

where the $x_k$ represent factors that cannot be measured or for which suitable data are not avail-

able, and e' represents that portion of Y that is distributed independently of $x_1 \ldots x_n$ and $x_k$. By omitting the $x_k$ which describe part of the "true" relationship, we find that we have not estimated the values of the $b_1 \ldots b_n$ that are needed, but instead the value of

$$b_1 + b_k b_{ik} \ldots\ldots + b_n + b_k b_{nk} \qquad [3]$$

where $b_{ik}$ is the regression coefficient of $x_k$ on $x_i$. Relation [3] describes a situation that is often referred to as "confounding." The fact that the relationships of most interest here are usually weaker than the relationships among independent variables can make the potential errors involved in [3] important.* However, note that according to [3], for a variable to confound, it must be correlated with both health outcome ($b_k$) and pollution ($b_{ik}$). Since most pollutants are correlated with one another, and usually to a degree stronger than their correlations with health, **any (unmeasured) pollutant with a known or suspected effect on health qualifies as a potential confounder with respect to the attempt to estimate effects specific to any other pollutant.** Also, any criteria pollutant must be deemed *a priori* to have the potential for adverse health effects.

Lack of independence of the $x_1$ is the rule rather than the exception in observational studies. Health is affected by weather and by pollution; pollution affects weather and vice versa, and there are exogenous factors (such as day-of-week activity and emission patterns) that affect both health and pollution. Given this state of affairs, the errors involved in selecting an

---

For example, in Los Angeles (Kinney and Ozkaynak, 1991) after adjustment for seasonality, the maximum correlation between any of four air pollutants and daily mortality was 0.13 while the correlations among the pollutants were in the range 0.43 to 0.88.

---

inappropriate model can be much larger than the statistical confidence limits derived from a particular model (Lipfert, 1994b), especially for large data sets. The prudent investigator will thus compare results obtained from a range of models, variables, and transformations for each new application.

Study Designs and Collinearity. The most common study designs involve either following a fixed population over time (the time-series study) or contrasting many populations over space at a fixed time (the cross-sectional study). In both designs, collinearity among pollutants is to be expected and has been observed. With time-series studies, all pollutants tend to respond to the same weather perturbations, since emissions tend to be steady over time (secondary pollutants such as ozone may be an exception). With cross-sectional studies, commonality of pollutants emitted by the same types of sources creates collinearity. In both cases, the use of a limited set of fixed air monitoring stations can involve sufficient uncertainties so as to mask the true degree of exposure collinearity and thus to make it impossible to determine the "responsible" pollutants with certainty. In all cases, the relationships among pollutants (and between pollutants and weather variables for time-series studies) are stronger than the relationships between pollution and health, which can make control of collinear variables crucial.

Effects of Measurement Errors. Actual (personal) exposures to air pollution are impractical to measure directly in populations large enough to have adequate statistical power in observational studies. The difference between measured concentrations and actual exposures may be expected to vary from day to day and by pollutant species. Some pollutants are distributed more uniformly throughout a community than others, and some penetrate into indoor environments more readily than others. Thus, we expect that errors in estimating exposure will vary by pol-

1. Confounding variables must be accounted for. If the model does not include all intercorrelated pollutants known to affect health, the regression coefficient for any one pollutant must be regarded as also including the effects of the omitted components of the pollutant mixture. This caution also applies to non-pollutant variables which may be correlated with pollution, such as ambient temperature.
2. For the pollutants for which measurements are available, comparisons among species must be made in such a way that the errors in exposure estimates are approximately equal. This requires a common basis for averaging time, for example. If data from a network of monitors are averaged, roughly the same number of stations should be used on all days and for all species.
3. The model must be appropriate to the physiological scenario. If peak exposures are thought to be more important than average exposures, nonlinear pollutant transformations should be investigated. If it is reasonable to expect a lag between exposure and response, a range of different lag structures should be investigated, including effects accumulated over the lag period and the possibilities of different lag structures for different pollutants, diseases, and age groups. If chronic effects are suspected, allowance must be made for the latency period required for development of the disease in question.
4. Depending on the outcomes of the preceding steps, an iterative procedure may be required to identify the most important pollutants and to estimate each of their separate contributions to the health outcome with an optimum model.
5. The regression model assumptions must be observed. This may include transforming the dependent variable to achieve normality and checking the regression residuals for heteroscedasticity.

We solve this relationship for the regression coefficients and their standard errors using ordinary least-squares methods. In the simulations performed, the *a priori* assumed values of $B_1$ and $B_2$ are unity. We assume that the pollutants $P_1$ and $P_2$ emanate from a common source or group of sources and thus are highly correlated in ambient air, even though they may not be measured at the same times and places. However, the relationships between monitored concentrations (the observed values in this case) and actual population exposure differ; additional (random) variance is present in the distribution of the values of $P_2$. We designate the unobservable value of this pollutant as $P_2$ (true) and the actual measurements as $P_2$ (measured). Thus $P_2$ meas. = $P_2$ true + measurement error ($e_m$) and our operational regression model is

$$D = B_0 + B_1P_1 + B_2'(P_2 + e_m) + e_D \qquad [5]$$

where $B_2'$ is the regression coefficient as determined in the presence of measurement error. In this sense, the measurement error term $e_m$ may be thought of as the differential sum ($P_2$ vs. $P_1$) of instrument and analytical error, spatial representativeness of the monitor network, effects of missing data, differences in lag structure of response, and differences in personal exposure relative to monitored concentrations. The magnitudes (but not the existence) of most of these effects are unknown.

Simulations were performed for 7 different combinations (referred to hereafter as "cases") of relative measurement error and assumed collinearity between $P_1$ and the true value of $P_2$, using 20 trials each. Each trial produces statistics for 5 different regressions, as discussed below. Crossplots are presented across regressions, with each trial constituting an "observation." Twenty trials were deemed sufficient to define average regression coefficients and *t* values, but are undoubtedly insufficient to completely define their distributions. The first of these simula-

which were defined as 1.0 in the formulation of the model; however, because of the collinearity, these coefficients are not significant and thus would likely be disregarded if they occurred with real data (in the absence of a priori information on the "true" model). In addition, the regression coefficients derived in the bivariate case (Regressions No. 1 and 2, one pollutant at a time) are inflated to essentially twice the expected value, since the effect of both pollutants is picked up by the single pollutant entered into the model. With measurement error, the coefficient for $P_2$(meas) is biased low with respect to that for $P_2$(true), as expected.

We see from Table 2 that the regression coefficient for $P_1$ is essentially unchanged from the bivariate case when regressed jointly with $P_2$ in the presence of measurement error (Regression No. 5), while the joint coefficient for $P_2$ has become nil, even though the basic underlying relationship assigns equal weight to both $P_1$ and $P_2$. Thus the "standard" multiple regression procedure yields an inflated coefficient for the pollutant with the lower measurement error and essentially zero for the pollutant with higher measurement error. These results confirm Cochran's earlier (1970) prediction, which was also based on numerical examples, that "interpretation of the $B_i$' as if they were $B_1$ can be quite misleading...." (typesetter: set Greek beta for B).

Figure 3 compares the *t* values for $P_1$ for the two joint regressions. When regressed jointly with the "true" $P_2$, $P_1$ loses significance, on average, because of collinearity. However, when $P_1$ is regressed jointly with $P_2$ in the presence of increased measurement error, $P_1$ loses less significance and its regression coefficients are virtually unchanged from the bivariate case (Figure 4). Note that all but one of the $P_1$ regression coefficients in Figure 4 exceeds the *a priori* or "true" value (1.0).

The corresponding *t*-value information for $P_2$ is given in Figure 5. In both cases, $P_2$ lost significance in most of the 20 trials. A plot of the joint regression coefficients for $P_1$ versus $P_2$

showed a negative relationship (Figure 6); a negative coefficient for one variable contributes to a more strongly positive value for the other.

Sensitivity Simulations. Additional simulation cases were performed in the same way to explore the robustness of these findings. These results (cases 2-6) are given in Table 3 along with the baseline result from Table 2 (case 1), in terms of the coefficients from Regression No. 5, the joint regression of mortality on $P_1$ and $P_2$(meas). There were little or no changes in the bivariate regression coefficients for these simulations.

Note that the regression coefficients tend to sum to the correct value of 2.0 in all 7 cases, but that the proper split (1:1) is approached only in simulation case 5. In this case, the introduced measurement error is small and the regression on $P_1$ and $P_2$(meas) is essentially the same as on $P_1$ and $P_2$(true). Note also that some of the simulation cases feature very modest amounts of pollutant collinearity and that biased coefficients still result from the joint regressions. Additional sensitivity runs (not shown) established that modest levels (relative to the error in $P_2$) of measurement error in $P_1$ had no effect on these findings.

## NUMERICAL EXPERIMENTS WITH 1973-80 PHILADELPHIA MORTALITY DATA

Data on mortality, weather, and air pollution (TSP, $SO_2$, and ozone) as measured in Philadelphia (Li and Roth, 1994) were used to further explore the ramifications of measurement errors in stepwise multiple regressions, but based on real instead of simulated data. The hypothesis here is that taking indoor/outdoor exposure relationships into account will add to the existing variance of the pollution terms, but in unknown ways. We created modified but non-

## WHAT IS THE CORRECT SHAPE OF THE DOSE-RESPONSE FUNCTION?

Most models employ linear specifications for the pollution terms; use of the logarithm of response in a log-linear model is tantamount to assuming an exponential form for the pollutant dose-response function. Table 4 presents bivariate correlations between six different pollutants and summer respiratory hospital admissions in Southern Ontario (Wyzga and Lipfert, 1993) for different functional forms. These results are for the entire year, adjusted for day of week and seasonality and lagged 3 days. The table shows that selection of the most "important" pollutant cannot be made independently of knowledge of the shape of the dose-response function. The differences in linear models are small, but we see that coefficient of haze (COH), $SO_2$, and $SO_4^{2-}$ show slightly higher associations when concave-downward functions are used (square root, log), while ozone and to a lesser extent TSP seem to point towards concave upward functions (square, exponential). The latter are more in keeping with the notion of pollution thresholds or log-linear models. $NO_2$ is seen to be indeterminant in this regard.

Table 5 compares beta coefficients (because they are nondimensional) for the logarithms of daily mortality in Philadelphia (1973-80), for the same functional forms. Data are shown for the 3 pollutants available in this study, taken jointly and separately, for 2 age groups. In general, functional form makes more difference in joint regressions than in separate regressions. We see some commonality with Table 4, in that $SO_2$ performs better with square root and log models (this was also the case over a much larger range of $SO_2$ values in the former East Germany [Spix et al., 1993]), and TSP performs better with the exponential and square functions. Ozone is seen to be indeterminate in joint regressions, but to indicate concave-upward functions in separate regressions.

We conclude from these data that it is necessary to know the appropriate form of a

dose-response function before the "best" pollutant can be selected with confidence.

## CONCLUDING DISCUSSION

This paper has reviewed the basic assumptions of multiple regression analysis, which are undoubtedly familiar to most readers and certainly to all practitioners. However, it is also common to overlook some of these concepts in the search to derive meaning from data. Certainly, peak-hour concentrations have been used with linear model specifications, nonlinear responses are rarely examined, and the effects of measurement error have been largely ignored.

We show by simulation and by numerical experiments that differences in the reliability of exposure estimates can be critical in the relationships between correlated variables in multiple (joint) regressions and thus in the identification of the "best" variables. Thus, epidemiological studies of community air pollution should recognize these inherent limitations in formulating their conclusions. Assignment of the total effect to the most statistically significant pollutant under such conditions appears to be fraught with the potential for error. Joint regressions of correlated variables will yield unbiased estimates of the "true" effects only when their relative levels of measurement (or exposure) error are similar. Further, even if the relative measurement errors are similar, if collinearity is severe, neither variable will appear to be significant even though their true effects are real. However, unless it can be shown (exogenously) that the effects of the omitted pollutant(s) are negligible, regressing each pollutant variable individually under these conditions will almost certainly overstate its effect. A research or pollution control program based on such findings would also appear to have a high potential for misdirection.

Finally, nonlinear transformations and the choice of pollutant metric can affect judgments about the relative importance of competing pollutants. While models with linear pollution

Chai, J.J. (1971), Correlated Measurement Errors and the Least Squares Estimators of the Regression Coefficient, J. Am. Stat. Ass. 66:478-83.

Cochran, W.G. (1970), Some Effects of Errors of Measurement on Multiple Correlation, J. Am. Stat. Ass. 65:22-34.

Cochran, W.G. (1972), Some Effects of Errors of Measurement on Linear Regression, in Proc. 6th Berkeley Symp. on Mathematical Statistics and Probability, ed. by J. Neyman, U. California Press, Berkeley, CA.

Fleiss, J.L., and Shrout, P.E. (1977), The Effects of Measurement Errors on Some Multivariate Procedures, Am.J.Public Health 67:1188-1191.

Kinney, P.L., and H. Ozkaynak, Associations of Daily Mortality and Air Pollution in Los Angeles County, Env.Res. 54:99-120 (1991).

Levine, R.V. (1990), The Pace of Life, Amer. Scientist 78:450-9.

Li, Y. and Roth, H.D. (1994), Daily mortality analysis by using different regression models in Philadelphia County, presented at the Colloquium on Particulate Air Pollution and Human Mortality and Morbidity, Irvine, CA, January 24-5, 1994.

Lipfert, F.W. (1993), A Critical Review of Studies of the Association Between Demands for

Hospital Services and Air Pollution, Environmental Health Perspectives Suppl. 101 (Suppl 2):229-268.

Lipfert, F.W. (1994a), Air Pollution and Community Health, Van Nostrand Reinhold, New York.

Lipfert, F.W. (1994b), Filter artifacts associated with particulate measurements: Recent evidence and effects on statistical relationship, submitted to Atm. Envir.

Lipfert, F.W., and Hammerstrom, T. (1992) Temporal patterns in air pollution and hospital admissions, Environmental Research 59:374-399.

Snedecor, G.W., and W.G. Cochran, Statistical Methods, 6th ed., The Iowa State University Press, Ames, Iowa. pp 164-166 (1967).

Spix, C., Heinrich, J., Dockery, D., Schwartz, J., Volksch, G., Schwinkowski, K., Coolen, C., and Wichmann, H.E. (1993), Air Pollution and Daily Mortality in Erfurt, East Germany, 1980-1989, Envir. Health Perspect. 101:518-27.

Theil, H. (1978), Introduction to Econometrics, Prentice-Hall, Englewood Cliffs, NJ. p. 137.

Wynder, E.L. (1987), Workshop on Guidelines to the Epidemiology of Weak Associations, Introduction. Preventive Medicine 16:139-141. See also Wynder, E.L. (1990), Epidemiological Issues in Weak Associations, Int.J.Epidemiology 19 (Suppl 1):S5-S7.

Wyzga, R.E., and Lipfert, F.W. 1993. Uncertainties in identifying responsible pollutants in environmental epidemiology. Presented at the 9th Congress on Aerosols in Medicine, Garmisch-Partenkirchen, Germany, April 1993.

Table 1 - Means and Standard Deviations (Baseline)

| Variable | Mean | Std. Dev. |
|---|---|---|
| D | 38.6 | 4.7 |
| $P_1$ | 6.1 | 0.93 |
| $P_2$ (true) | 7.6 | 0.94 |
| $P_2$ (meas) | 7.6 | 1.18 |

**Table 2 Baseline Regression Statistics**

| | separate | | | joint | |
|---|---|---|---|---|---|
| Regress. | # 1 | 2 | 3 | 4 | 5 |
| Statistic | D vs $P_1$ | D vs $P_2$true | D vs $P_2$meas | D vs $P_{1,2}$true | D vs $P_{1,2}$meas |
| $B_1$ | 1.88 | | | 1.06 | 1.88 |
| sigma$B_1$ | 0.45 | | | 2.1 | 0.78 |
| t | 4.47 | | | 0.6 | 2.94 |
| $B_2$ | | 1.81 | 1.07 | 0.85 | -0.01 |
| sigma$B_2$ | | 0.46 | 0.27 | 2.1 | 0.5 |
| t | | 4.46 | 3.21 | 0.48 | 0.0 |
| $R^2$ | 0.17 | 0.17 | 0.098 | 0.17 | 0.185 |
| Std error of est. | 4.18 | 4.18 | 4.37 | 4.18 | 4.17 |

note to typesetter: set Greek sigma for "sigma"

**Table 3 Summary of Simulation Results for the Joint Regression of Mortality on $P_1$ and $P_2$meas.**

| Case # | Pollutant Correlations | | Average Regression Coefficients | |
|---|---|---|---|---|
| | $P_1$,$P_2$true | $P_1$,$P_2$meas | $P_1$ | $P_2$meas |
| 1. | 0.96 | 0.75 | 1.88 | -0.01 |
| 2. | 0.84 | 0.75 | 1.71 | 0.33 |
| 3. | 0.90 | 0.84 | 1.75 | 0.36 |
| 4. | 0.96 | 0.91 | 2.04 | -0.08 |
| 5. | 0.72 | 0.65 | 1.08 | 0.87 |
| 6. | 0.66 | 0.29 | 1.76 | 0.18 |
| 7. | 0.50 | 0.27 | 1.52 | 0.56 |

**Table 4 - Bivariate Correlations of Respiratory Hospital Admissions in Southern Ontario, 1979-85, Adjusted for Season and day of Week, with Various Air Pollutants and Functional Forms of the Pollutant Dose-Response Function**

| | Function | | | | |
|---|---|---|---|---|---|
| Pollutant | linear | exponential | square | square root | natural log |
| $NO_2$ | 0.038 | 0.024 | 0.038 | 0.037 | 0.033 |
| COH | 0.034 | 0.032 | 0.030 | 0.036 | 0.037 |
| $SO_4^=$ | 0.034 | 0.027 | 0.029 | 0.035 | 0.033 |
| $O_3$ | 0.029 | 0.053 | 0.042 | 0.020 | 0.008 |
| $SO_2$ | 0.026 | 0.011 | 0.017 | 0.029 | 0.032 |
| TSP | 0.021 | 0.018 | 0.020 | 0.018 | 0.012 |

Table 5 - Beta Regression Coefficients for Philadelphia (Logarithm) Mortality and Air Pollution in Joint and Separate Regressions for Various Functional Forms (n=2459).

A. Deaths at Ages < 65

| | | LINEAR | EXP | SQUARE | SQ ROOT | LOG |
|---|---|---|---|---|---|---|
| joint regr | TSP | 0.0375 | *0.0509* | *0.0525* | 0.0311 | 0.0359 |
| | $SO_2$ | *0.049* | 0.0417 | 0.0389 | *0.0551* | *0.0522* |
| | $O_3$ | 0.022 | 0.0182 | 0.012 | 0.0317 | 0.022 |
| sep regr | TSP | 0.068 | *0.075* | *0.074* | 0.065 | 0.062 |
| | $SO_2$ | *0.071* | 0.071 | 0.066 | *0.072* | *0.07* |
| | $O_3$ | 0.021 | 0.021 | 0.021 | 0.019 | 0.005 |

B. Deaths at Ages 65+

| | | LINEAR | EXP | SQUARE | SQ ROOT | LOG |
|---|---|---|---|---|---|---|
| joint regr | TSP | 0.0736 | *0.0832* | *0.0967* | 0.0629 | 0.0566 |
| | $SO_2$ | *0.0835* | 0.0731 | 0.0493 | *0.0991* | *0.104* |
| | $O_3$ | 0.0688 | 0.0684 | 0.0619 | 0.0704 | 0.0795 |
| sep regr | TSP | 0.133 | *0.134* | *0.134* | 0.128 | 0.0121 |
| | $SO_2$ | *0.14* | *0.135* | 0.117 | *0.145* | *0.14* |
| | $O_3$ | 0.077 | 0.082 | 0.086 | 0.053 | 0.05 |

Covariates include date, dew point, hot days, mean temperature, and change in barometric pressure.

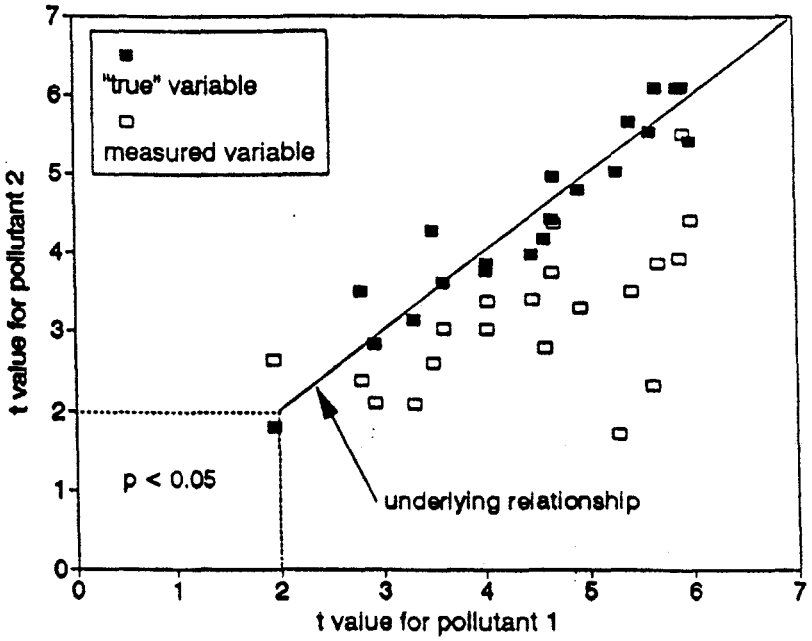


Figure 2. Relationships among the coefficient t values for the 20 trials of the baseline simulation, bivariate regressions. T values from regressions 2 and 3 are plotted against the t value for regression 1, for each trial. Source: Lipfert, 1994a.

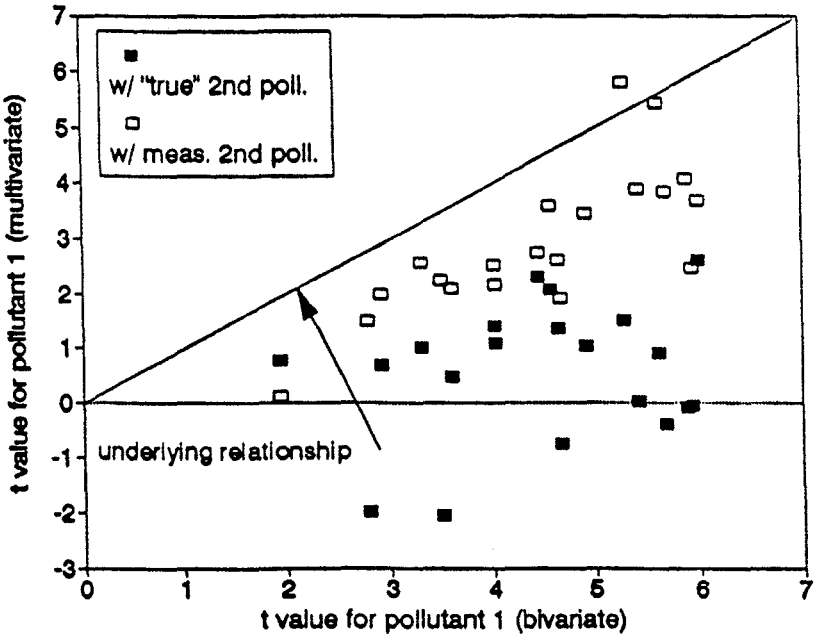


Figure 3. Relationships among the $P_1$ t values for the 20 trials of the baseline simulation. T values from regressions 4 and 5 are plotted against the t value for regression 1, for each trial Source: Lipfert, 1994a.

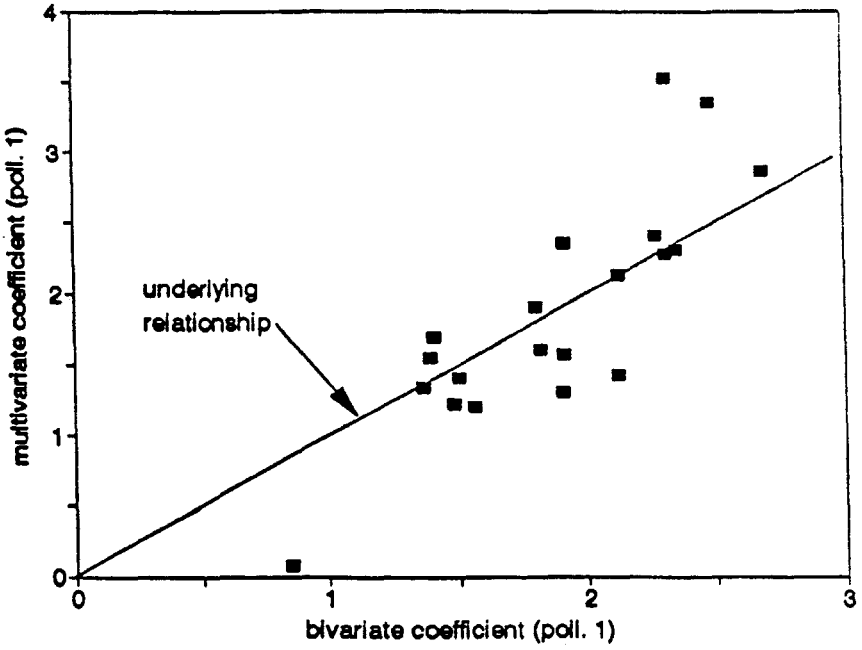


Figure 4. Relationships among the regression coefficients for $P_1$ for the baseline simulation. Coefficients from regression 5 (joint with $P_2$(meas) are plotted against the (bivariate) coefficient value for regression 1, for each trial. Source: Lipfert, 1994a.

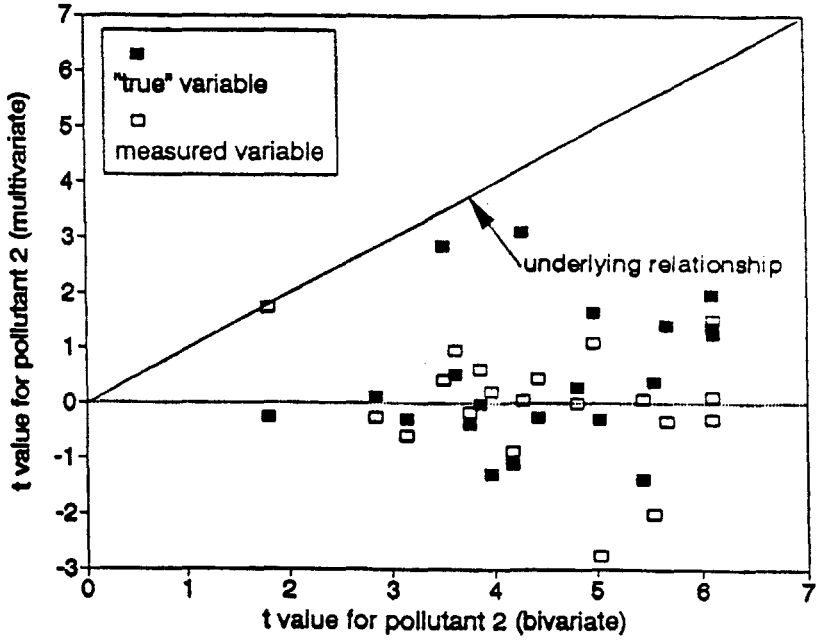


Figure 5. Relationships among the $P_2$ t values for the 20 trials of the baseline simulation. T values from regressions 4 and 5 are plotted against the t value for regression 2, for each trial. Source: Lipfert, 1994a.

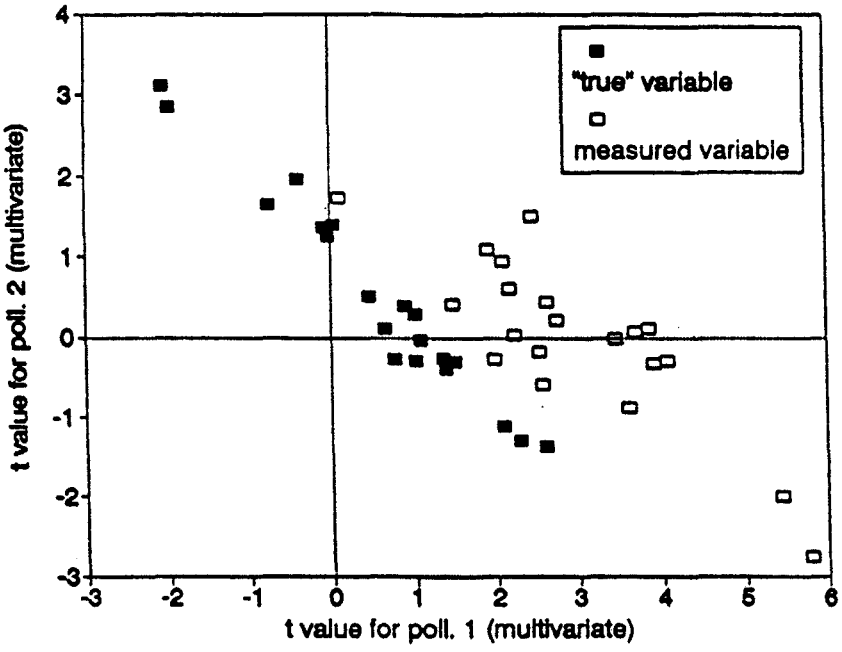


Figure 6. Relationships among the coefficient t values for the 20 trials of the baseline simulation. T values for $P_1$ from regressions 4 and 5 are plotted against the t value for for $P_2$, for each trial. Source: Lipfert, 1994a.