Environmental Health Perspectives
Vol. 89, pp. 159-168, 1990

The Relationship of Daily Mortality to Suspended Particulates in Santa Clara County, 1980–1986

by David Fairley*

This paper explores the relationship between daily mortality and suspended particulates in Santa Clara County, CA, for years 1980 to 1986. An association was found between high particulate concentrations and increased mortality. This association persists after adjustment for temperature, relative humidity, year, and seasonality. Contrary to expectation, the magnitude of the particulate effect appears the same or larger than that estimated for London, despite Santa Clara County's cleaner air. The persistence of an effect at these lower particulate concentrations suggests that the particulate variable may be acting as a surrogate for some constituent particles, such as acid aerosols.

Introduction

In the 1950s, the citizens of London were subjected to periods of very high particulate concentrations that resulted in daily deaths far above the norm. As a result, the first efforts were made to regulate particulate levels, banning coal burning in certain areas. Since then a number of studies [e.g., Mazumdar et al. (*1*), Ostro (*2*), Schwartz and Marcus (*3*) for London; Glasser and Greenburg (*4*), Schimmel and Greenburg (*5*), Schimmel and Murawski (*6*) for New York City] have shown that increased deaths are associated with increased particulate levels, even when the levels are much lower than those of London in the 1950s. These studies used regression analysis to estimate rough dose-response relationships and have served as key elements in establishing U.S. standards for maximum allowable particulate concentrations.

While the evidence for serious health effects from particulates is now well established, the actual cause-and-effect mechanism is not well understood. It is known that almost all large particles (pollen and some household dust, for example) are filtered out in the nose and throat. Only particles less than about 10 μm get into the lungs, and only the so-called respirable particles, those less than about 3 or 4 μm, can penetrate deep into the lungs. Taking this into account led to setting the U.S. standards in terms of PM_{10}, the weight, per cubic meter, of airborne particles less than 10 μm.

There remains the additional question of which particles are actually doing the damage. It is likely that sheer mass is a factor, but it is also reasonable to suspect that certain components (sulfates or nitrates, perhaps) may be considerably more damaging per unit weight than, say, dust particles. Different areas have different mixes of particulates. Thus, it is unclear whether the same particulate/mortality relationships found in London or New York City, which both have had high sulfate levels, would apply for Santa Clara County, which does not.

This analysis originated because of the Santa Clara Criteria Air Pollutant Benefit Analysis (*7*) study, which applied the regression coefficient from a London study to the Santa Clara County population. It concluded that 40 deaths per year might be avoided if the particulate levels were reduced to the State standard. The present study examines whether country-specific data support this conclusion.

Santa Clara has a more moderate climate, lower particulate levels, and a lower mortality rate than London. In addition, United Kingdom residents appear to be more susceptible to respiratory problems than U.S. residents. The age-adjusted death rate attributed to respiratory causes was 23.6 per 100,000 for England and Wales for 1984 compared with only 8.6 in the U.S. (*8*), although some of this difference is probably due to differences in standard medical practice in assigning cause of death in the two countries. Thus, it was expected that if any mortality/particulate relationship were found at all for Santa Clara County it would be smaller than that found for London.

Data

Mortality data were obtained from the Santa Clara County Health Department for the years 1980 to 1982 and 1984 to 1986 (1983 was missing). The data came in the form of individual records on a computer tape. Each record included age, sex, death date, census tract, whether the person died in or out of county, location of death (hospital, rest home, at home), and cause of death. To parallel other studies, the data were reduced to total number of deaths per day. A slight

*Bay Area Air Quality Management District, 939 Ellis Street, San Francisco, CA 94109.

modification was made of limiting the totals to county residents who died in county of nonaccidental causes. In the remainder of this study, this variable will be referred to as "mortality."

The total number of deaths per year was about 9000, while the mortality was about 6800, or about 25 and 18 deaths per day, respectively. This rate is considerably smaller than the approximately 300 per day for London and New York City. Essentially, this means a smaller sample size, and hence less power to detect the effect of particulates. However, it may have simplified the analysis by almost eliminating any autocorrelation in the mortality time series.

There are several methods currently used to measure particulate levels. Because the particulate standards are written in terms of PM_{10}, it would have been desirable to use this variable in the analysis. However, these measurements have been made in the San Francisco Bay Area only since 1985, and there is only one measurement taken every 6 days. So using PM_{10} would have resulted in a major decrease in the amount of data that could be used. Instead, coefficient of haze (COH) was chosen because it was available on a daily basis from the Bay Area Air Quality Management District monitor at San Jose for the entire period under study. How the analysis was affected by the use of particulate measurements at one site instead of some more accurate measure of population exposure will be taken up in the "Discussion."

Table 1. Yearly means.

	1980	1981	1982	1983	1984	1985
Daily mortality	17.7	18.1	18.3	18.8	19.5	18.8
San Jose COH	67.9	56.3	59.5	60.9	86.8	70.0

The COH instrument, called an AISI sampler, operates by drawing air through a section of filter paper for a 2-hr period so that airborne particles are deposited on the paper. A light is shone through the paper both before and after the air is drawn through, and the amount of light transmitted is measured by a photocell. The COH measurement is based on the ratio of the current produced by the photocell prior to and after sampling.

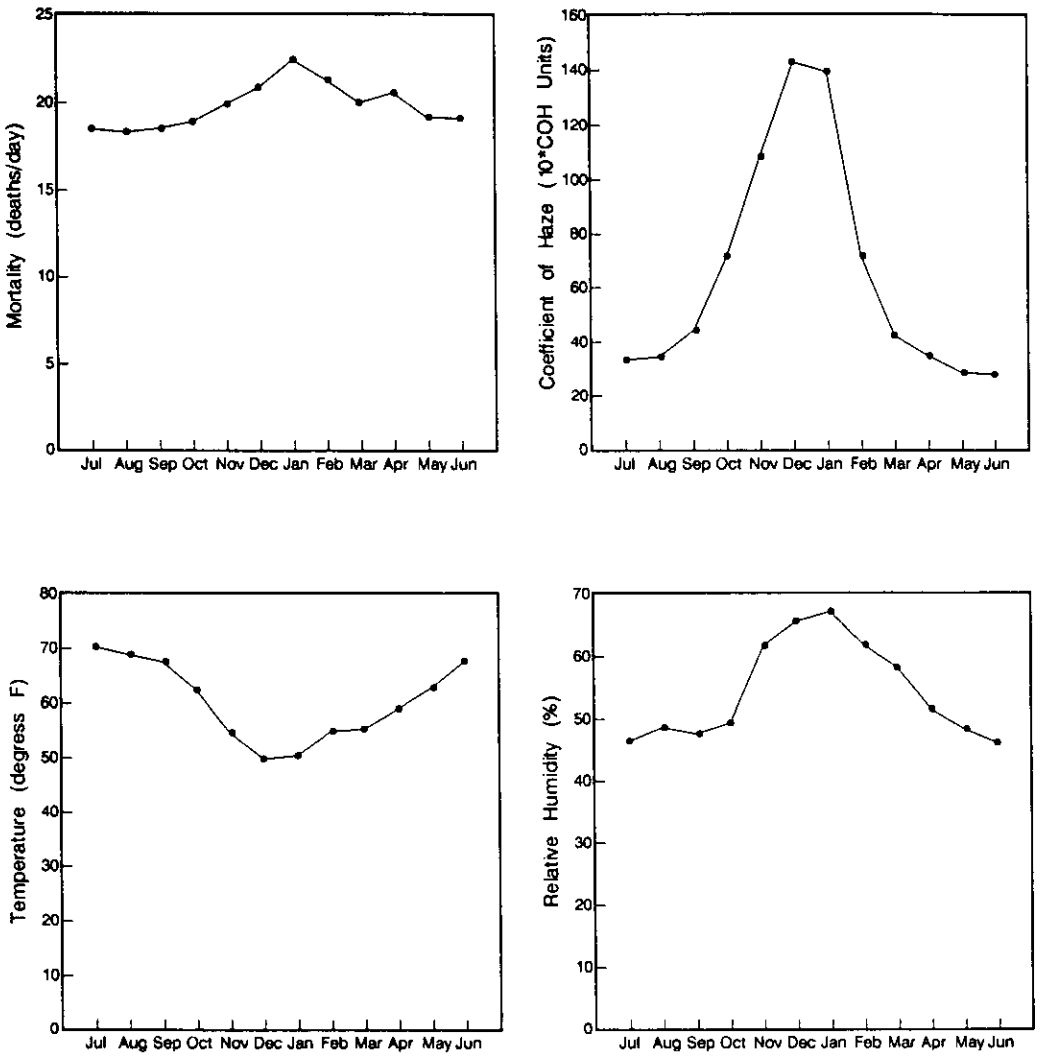

FIGURE 1. Seasonal variation in mortality and the independent variables.

For reasons discussed below, only the months of November, December, and January were used in this study. The median ratios of COH to PM_{10} during these months in 1985 and 1986 were 1.87 and 1.64, respectively, and the correlations were 0.79 and 0.81. Thus, there is a relatively high correlation between COH and the particulate measure upon which the national particulate standard is based, PM_{10}.

The major sources of particulates in Santa Clara County are re-entrained road dust (around 60%) and construction dust (around 20%)(*9*). Re-entrained road dust contains, in addition to soil particles, engine oil which includes various metals; tire particles; and sulfates. Other particles are formed in the atmosphere, notably nitrates and sulfates. Based on gravimetric sampling, nitrate make up about 7% of all particulates smaller than 10 μm, and sulfates about 6% by weight (*10*). The other variables used in this analysis were daily mean temperature and relative humidity at 4 P.M., but measured at the San Jose City Hall.

Summary Statistics

Table 1 shows the yearly means for mortality and COH. There is a general increase in daily death with year for Santa Clara County, perhaps due to increasing population. There is no strong trend for COH, but 1985 was unusually high due to nearby construction and meteorology.

Figure 1 shows the monthly means for all four variables during the period of study. All show seasonal patterns, with COH being particularly striking. The months November, December, and January, hereinafter called the winter months, have COH levels three to four times those of the summer months.

For the following reasons, the analysis was limited to the winter months: *a*) It seems likely that the mortality/particulate relationship would be different from the summer months, making regression analysis difficult. *b*) It is very possible that it is only accidental that the pattern of high winter/low summer mortality matches the seasonal pattern

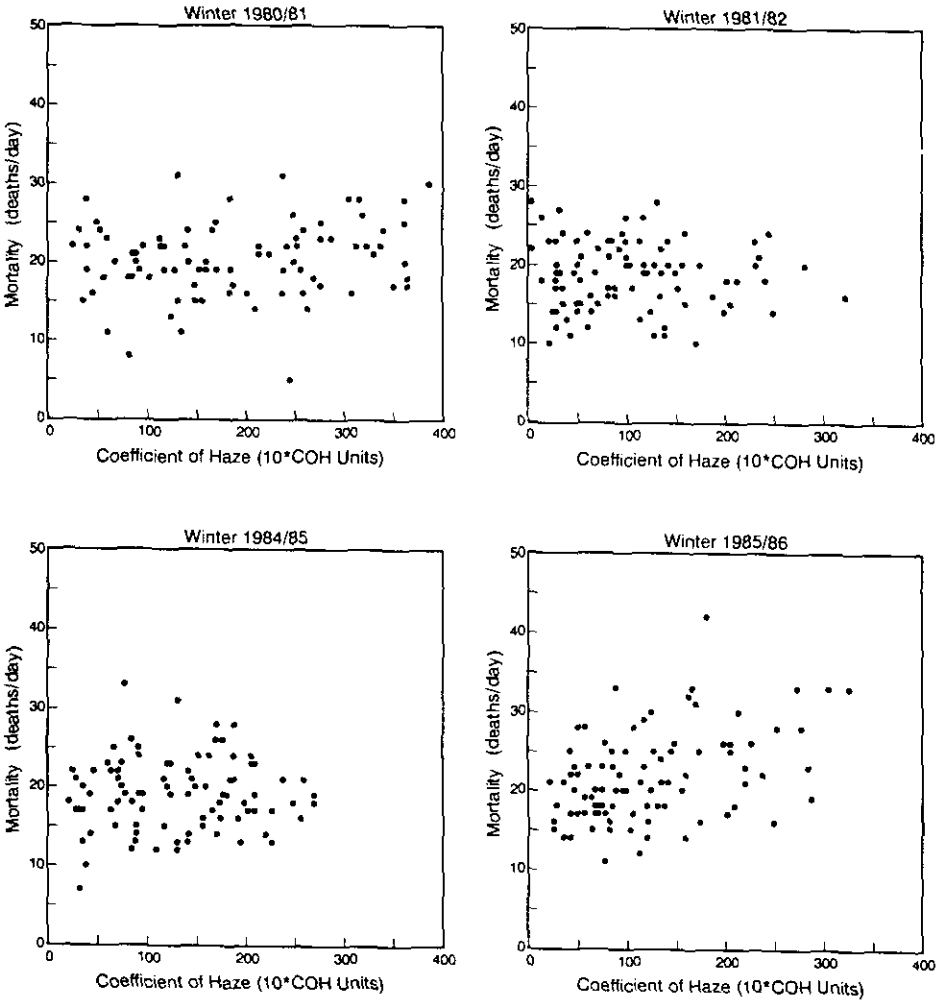

FIGURE 2. Daily mortality versus COH for Santa Clara County for the four complete winters 1980/81, 1981/82, 1984/85, and 1985/86.

Table 2. Autocorrelations of daily mortality.

	1	2	3	4	5	6	7	8	9	10
Autocorrelation	0.14	0.09	0.10	0.14	0.10	0.06	0.15	0.05	0.16	0.00

of the other variables in Figure 1. *c*) Because COH is so much higher in the winter, any COH/mortality relationship should show up most clearly in the winter months. *d*) The London studies only considered the winter months November through February.

Regression Analysis

Previous studies have used a variety of models and variable transformations. Figure 2 shows daily mortality versus same-day COH for the four complete winters, 1980/81, 1981/82, 1984/85, and 1985/86. The plots show no clear indication of nonconstant variance, skewness in either variable, or serious outliers. For these reasons transformation or weighted least squares were not used. (The mortality variable can be considered approximately Poisson, so a square root transformation might have been considered. See "Poisson Regression" for some discussion.) Questions of nonlinearity are dealt with by fitting second-order models.

Perhaps because of the lower death count there is little autocorrelation in the mortality time series. Table 2 shows the first 10 mortality autocorrelation coefficients. These were obtained by computing the autocorrelations for each of the four complete winters and averaging them across years. The standard error is $1/[4(92)]^{1/2} \cong 0.05$. While several are significant, none is larger than 0.2. Performing the regressions reduces the autocorrelations further. Thus, autocorrelation should have only a minor effect on significance levels.

The predictors were chosen on the basis of previous studies. Both the New York and London studies used average daily temperature, humidity, and some measure of particulate levels. Many of these studies also included sulfur dioxide, but as mentioned earlier, this is not a serious pollutant in Santa Clara County and was not included in this analysis.

The effects on health of several days of high particulate levels may be cumulative. Thus, regressions were done both with same-day COH and temperature, and also with lags of these variable, i.e., using previous days' COH and temperature to predict today's mortality. To account for the increasing death rate, indicator variables for year were included in most regressions. Any remaining temporal effect in the mortality data was accounted for by a third-order polynomial in day (see the following model equation).

A subtle question is whether a coinciding period of high mortality and high particulates can be considered causal. Did a period of high particulates cause the corresponding period of high mortality? The conservative approach is to filter out these lower frequency associations because these associations might be accidental. One method used in previous studies was to transform both the dependent and independent variables to deviations from a 15-day moving average. The method used in this study was to fit separate third-order polynomials to each year. The effect on the COH coefficients of these two methods appears similar. The general model fit was

$$Y_{it} = \beta_0 + \mu\,\beta_{1k}C_{i,t-k} + \mu\,\beta_{2k}T_{i,t-k} + \beta_3 H_{it} + \gamma_{i0} + \gamma_{i1}(t\text{-}46.5) + \gamma_{i2}(t\text{-}46.5)^2 + \gamma_{i3}(t\text{-}46.5)^3 + \epsilon_{it}$$

with
- i = winter (i=1,2...,8),
- t = day of the winter (t=1,2,...,92),
- k = lag (k=0,1,...,6),
- y_{it} = mortality on day t in winter i,
- C_{it} = COH,
- T_{it} = mean temperature,
- H_{it} = 4 P.M. relative humidity.

Second-order models including $C_{i,t}^2$, $T_{i,t}^2$, and $C_{i,t}T_{i,t}$ were also fit.

Table 3 presents the results of these analyses. There were a total of 549 observations: 6 Novembers, Decembers, and Januaries comprising 6×(30+31+31)=552 days with 3 missing. However, there were 8 winters beginning with January 1980 and ending with November/December 1986. The COH coefficient represents the change in expected daily deaths per change of one COH unit (all else held constant). In Table 3, the F column represents the F-statistics comparing the model with the largest submodel above it, and the p-value column contains the p-values corresponding to the F-statistics. This shows whether the new variables improve the fit of the models significantly.

In model 1, only COH is fit. The COH coefficient is statistically significant, but this model does not account for any potentially confounding factors. Comparison of models 1 and 2 shows that year effects are highly significant.

The inclusion of temperature and humidity (model 3) causes a marginal improvement in goodness of fit (higher adjusted R^2). Because of this and to parallel previous studies as much as possible, these variables were included in the succeeding models even though they are not statistically significant. On the other hand, the addition of second-order terms—the squares of COH and temperature, and the cross-product of COH and temperature—results in no improvement in goodness of fit. Thus, there is a negligible amount of nonlinearity in the relationship of these variables to mortality, so the remaining models omit these variables.

The addition of a winter seasonal curve in model 4 produces a highly significant improvement over model 3, showing that there is a consistent pattern in winter mortality unexplained by COH. However, there is little change in the COH coefficient, which indicates that COH was not just a surrogate for a seasonal pattern in mortality.

Fitting separate trend terms for each winter (model 5) yields a significant improvement in fit. Thus, there are year-specific trends not explained by same-day COH. In contrast to model 4, the COH coefficient is now about 25% smaller, so that these trend terms can explain some of the mortality variation previously attributed to COH. Fitting separate seasonal curves to each winter (model 6) does not improve the fit, but it deflates the COH coefficient much further. The COH coefficient from this model is similar to the coefficient found by computing deviations from a 15-day moving average (model 7). The effects of fitting separate

Table 3. Mortality regressions.

Model no.	Description (n = 549)	No. of predictors	COH[a] coefficient, ×100	SE, ×100	Adjusted R^2, %	F-statistic[b]	p-Value
1	COH only	1	0.99	0.25	2.7	15.99	<0.0001
2	Model 1+year intercepts[c]	8	0.94	0.25	6.7	4.44	<0.0001
3	Model 2+weather[d]	10	0.83	0.29	7.1	1.97	0.16
4	Model 3+season[e]	13	0.84	0.29	11.3	10.96	<0.0001
5	Model 4+year/day intercepts[f]	20	0.64	0.30	13.2	2.72	0.01
5a	Model 5−COH	19	—	—	12.2	—	—
6	Model 5+year/season[g]	34	0.48	0.32	13.4	1.10	0.34
6a	Model 6−COH	33	—	—	13.0	1.35	0.18
7	15 day deviations[h]	3	0.45	0.29	0.3	—	
8	Model 4+6 temperature lags[i]	19	0.61	0.31	11.2	1.15	0.33
9	Model 4+2 COH lags[i]	15	0.29	0.41	12.8	3.74	0.03
10	Model 4+6 COH lags[i]	19	0.20	0.50	12.9	1.65	—

[a] COH, coefficient of haze.

[b] F-values represent the improvement in fit over the next smaller model. The F-value for model 1 is the improvement over a constant regression. The F-value in 6a is the improvement in a fit of model 6a over model 5a.

[c] Separate intercepts for each of eight winters.

[d] Daily mean temperature and 4 P.M. humidity at San Jose.

[e] Third-order polynomial in day, t, $t=1,2,\ldots,92$.

[f] Separate linear (t) terms fit to each year.

[g] Separate third-order polynomials fit to each year.

[h] Each winter's NOAC mortality is standardized by dividing by the winter mean. Fifteen-day moving averages are computed for this standardized mortality and then subtracted from the daily values, forming a sequence of deviations, e.g., the average standardized mortality for January 1 through 15 would be subtracted from the January 8 mortality. Similarly, deviations are computed for COH, temperature, and humidity. A linear regression of the mortality deviations on these COH and weather deviations is then performed. To get the COH coefficient and SE above, the regression coefficient and SE were multiplied by the average winter mean of 19.85 deaths/day.

[i] Lagged temperature and COH are previous days' values of these variables.

seasonal curves and subtracting a 15-day moving average are similar: each acts to filter out much of the slowly varying component in mortality time series. The decrease in the COH coefficient may indicate that mortality has an irregular, slowly varying component that coincides with a similar such component in COH, so that filtering out this component would lower the correlation between COH and mortality. However, if models 5 and 6 are fit without COH (models 5a and 6a), there is no significant improvement in fitting separate seasonal curves. Thus, there is not strong evidence of an effect from a year-specific slowly varying factor independent of COH.

When temperature lags are included (model 8), there is no improvement in fit over model 4, but some decrease in the COH coefficient. This may be due to temperature also sharing a slowly varying component with COH.

Adding lagged COH coefficients does result in a significant improvements in fit for lags 1 and 2 (model 9). The estimated COH coefficients for lags 0, 1, and 2 were 0.0029, 0.0035, and 0.0055, respectively. Adding further lags does not add significantly to model 9 nor to model 4 (p-values > 0.50 and > 0.05, respectively). These results appear to indicate that the effects of COH are cumulative or delayed by 1 or 2 days. The estimated effect of today's COH is greatly reduced, suggesting that today's COH is a surrogate for a better predictor, such as a weighted average of recent COH, perhaps.

Model Assumptions and Bounds on Explainable Variation

As previously discussed, there do not appear to be serious problems with nonconstant variance or outliers or skewness in the predictors. A look at the residuals from model 4 shows several large outliers: 4 above 3 regression SDs. The largest was an SD of 4.03, on a day when 42 people died. The prediction for that day was 23.9. The likelihood of a Poisson random variable with mean 23.9 being greater than or equal to 42 is 0.0005. The chance of seeing at least 1 day out of 549 that much above its mean is $1 - 0.9995^{549} = 0.24$. Thus, the outlier does not appear exceptionally improbable. The effect on the regression of these outliers is minimal. The first-order autocorrelation of the residuals for model 4 was 0.04, which is not significantly positive. Thus, none of the usual assumptions of linear regression appears seriously violated.

The R^2 values in Table 3 are all quite low both in absolute terms and also relative to those found in other studies. However, a large fraction of the mortality variation may not be explainable without introducing predictors aimed at subsets of the population. It is arguable that the daily numbers of deaths act as a sequence of Poisson random variables, which are, for all practical purposes, independent: The probability that a randomly selected individual will die today is quite low—we see around 20 deaths per day only because there are over a million people in Santa Clara County. Thus, the number of nonaccidental deaths today is a collection of rare, nonsimultaneous, and almost independent events. These are the criteria for a sequence of independent Poisson random variables.

Under the Poisson assumption, the daily mean represents a lower bound for the regression mean squared error (MSE). The sample mean and variance of the 8 winters of mortality data were 19.85 and 23.20, respectively. If the mean value represents a real lower bound, then the larger models in Table 4 approach the limit of the amount of explainable varia-

Table 4. Yearly regressions.

Winter	Same-day COH[a]			Lag 2 COH		
	Coefficient	SE	t	Coefficient	SE	t
1980/81	0.0118	0.0058	2.00	0.0121	0.0052	2.33
1981/82	0.0037	0.0076	0.48	0.0074	0.0066	1.12
1984/85	0.0149	0.0094	1.59	0.0209	0.0088	2.37
1985/86	0.0041	0.0077	0.54	0.0093	0.0067	1.38

[a]COH, coefficient of haze.

tion (i.e., explainable by predictors that are constant across the population). For example, model 4 has an MSE of 20.58. This represents (23.20 − 20.58)/(23.20 − 19.85) × 100% = 78% of the explainable variation. Model 5 has an MSE of 20.13, which is 92% of the explainable variation. If this reasoning is correct, there is little left to explain with global variables. To improve the predictions would require the use of variables that are specific to subsets of the population (e.g., COH levels and temperatures in various neighborhoods, age, sex, or individual case histories).

Regressions for Individual Years

Table 4 presents regressions for the four complete winters. The predictors for the left-hand set of regressions are same-day COH, temperature, humidity, and a seasonal polynomial. The right-hand set is the same as the left, only with COH 2 days previous replacing same-day COH.

The regression coefficients can be compared using the studentized range distribution. Since the coefficients have slightly different SDs, the minimum of the SD provides a bound. The studentized range statistic is (0.0149−0.0037)/0.0058 = 1.93, which is less than 3.24, the 90th percentile of the studentized range distribution for 4 samples with an infinite number of degrees of freedom. Thus, the regression appear consistent in the sense that the estimated COH coefficients for the four winters are not significantly different from each other. Although most of the coefficients are not significant, they ar all positive, and even the smallest, 0.0037, represents a sizeable effect (see "Discussion").

The lag 2 coefficients are consistently larger than same-day COH, while the SEs are similar. This may indicate a cumulative or delayed health effect from elevated particulate levels.

Regressions by Death Classification and by Age

If particulates are influencing mortality and not just a surrogate for some other factor, then one would expect a higher death rate among those who are most sensitive to particulate effects. Specifically, particulates are thought to cause breathing and lung difficulties, so it might be expected to affect those with respiratory problems most seriously. Only the primary cause of death is listed on the death certificate, so in some cases respiratory problems might precipitate the death, but the death is attributed to another cause. In other words, it would not be unexpected to see a positive correlation among other nonaccidental death categories besides respiratory.

Table 5. Deaths in various disease categories.

Cause of death	Average daily deaths, full year	Percent of total	Average daily deaths, winter	Percent of total	Percent winter increase
Nonaccidental	18.26	91.9	19.59	92.9	7.2
Respiratory[a]	1.62	8.1	1.80	8.5	11.1
Cancer[b]	4.79	24.1	4.84	22.8	1.0
Circulatory[c]	8.82	44.3	9.65	45.5	9.4
Other[d]	3.03	15.2	3.30	15.6	8.9
Accidental[e]	1.62	8.1	1.50	7.1	−7.4
Total	19.88	100	21.69	100	6.0

[a]International Classification of Disease (ICD) categories, 11, 35, 472–519, 710.0, 710.2, 710.4.
[b]ICD categories, 140–209.
[c]ICD categories, 390–459.
[d]All other ICD categories less than 800.
[e]ICD categories 800–999.

The data were divided into major causes of death according to the International Classification of Diseases (ICD). Summary statistics are provided in Table 5. The category of the circulatory system (numbers 390–459 in the ICD) includes myocardial infarction, arteriosclerosis, and stroke and constitutes almost 50% of all deaths (Table 5). A second category is cancers (numbers 140–209), constituting roughly 25% of all deaths. A third category is respiratory diseases (categories 472–519, 11, 135, 277,.3, 710.0, 710.2, and 710.4), which includes emphysema, chronic bronchitis, pneumonia, and tuberculosis and constitutes just under 10% of all deaths. The "other category" includes all other nonaccidental deaths, those not induced by injury or poisoning (all other categories below 800). These include infant deaths and cirrhosis of the liver. The accidental death category is all numbers 800 through 999, including traffic deaths, homicides, and other types of accidents.

Table 5 shows that the total death rate in the winter months is 6% higher than the year as a whole, but that accidental deaths actually decline somewhat. So there is a noticeable increase in the number of nonaccidental deaths. There is essentially no change in the number of cancer deaths, and deaths due to other causes rise about 10%.

Table 6 presents the results of models 4 and 5 with the dependent variable being daily deaths due to the various categories of disease. Respiratory deaths are highly significant for both models. Nevertheless, this coefficient is smaller than the coefficient for circulatory deaths because respiratory deaths constitute a smaller fraction of all deaths. The last column of the table attempts to put the coefficients on an equal footing. It represents the percentage of daily deaths in a particular category that are affected by a change of 1 COH unit. For example, there is an expected decrease of 0.172% in daily respiratory deaths with a decrease of 1 COH unit. The percentage of respiratory deaths affected is about four times the number from other categories. It should also be noted that, although they are at best borderline significant, the estimated effect of COH on both cancer deaths and circulatory deaths is positive. This is in line with the findings of a similar study for New York City (*5*), where positive, and in some cases significant, results were obtained with coronary heart disease, hypertensive heart disease,

Table 6. Regressions by cause of death.

Dependent variable	Regression coefficient, ×100	SE	t-Value	p-Value	COH rate per death, %
Model 4					
Nonaccidental	0.84	0.29	2.90	0.002	0.043
Respiratory	0.31	0.09	3.63	<0.001	0.172
Cancer	0.21	0.14	1.58	0.114	0.043
Circulatory	0.40	0.19	2.11	0.035	0.041
Other	−0.09	0.13	−0.73	0.468	−0.027
Accidental	0.01	0.07	−0.10	0.872	−0.007
Model 5					
Nonaccidental	0.48	0.32	1.50	0.12	0.024
Respiratory	0.26	0.09	2.86	0.004	0.144
Cancer	0.18	0.14	1.27	0.203	0.037
Circulatory	0.30	0.20	1.52	0.128	0.031
Other	−0.09	0.13	−0.72	0.474	−0.027
Accidental	−0.03	0.08	−0.39	0.698	−0.020

and cancer of the respiratory system.

A finer breakdown by individual diseases lead to conflicting results. Of the 20 leading causes of death (i.e., the individual ICD numbers that had the highest number of winter deaths), pneumonia (486), colon cancer (153), and strokes (436) were the only statistically significant causes based on model 5. The categories of respiratory disease excluding pneumonia and cerebrovascular disease excluding stroke (436) were also significant.

Regressions were also done on nonaccidental deaths by age at death. The data were divided into those 70 and older at death and those younger than 70. The daily mean numbers of deaths were 12.17 and 7.68 for ≥ 70 and < 70, respectively. The COH coefficients (SEs) from model 5 were 0.0051(0.0022) and 0.0014(0.0019), respectively, the coefficient for the ≥ 70 category being borderline significant, while the < 70 category was insignificant. Comparison of the death rates shows a 0.0051/12.17 = 0.0419% rate for the ≥ 70 group compared with a 0.00147/7.68 = 0.0182% rate for the <70 group. Although the former rate is more than double the latter, suggesting that elderly people maybe more strongly affected by COH (or what is a surrogate for), the difference in rates is not statistically significant.

Poisson Regression

The mortality time series is a sequence of approximately independent Poisson random variables, so it seemed reasonable to try Poisson regression (Table 7). The results are quite similar to those found in Table 3. The largest models fit were similar to models 4, 5, and 6 of Table 3. These large models fit the data adequately in the sense that the unexplained variation could be due to chance. (See "Appendix" for details.)

Discussion

These analyses show that in Santa Clara County, mortality tends to increase on days with increased particulate levels. Moreover, contrary to expectation, the estimated magnitude

Table 7. Poisson regressions.

Model no.	Description (n = 549)	p	COH coefficient[a]	−2log(likelihood ratio) vs. submodel[b]	−2log(likelihood ratio) vs. saturated[c]
1	COH only	1	0.0097	—	620.2
			(<0.0001)[d]	—	(0.02)
2	(1) + years intervals	8	0.0093	33.58	586.6
			(<0.0001)	(<0.0001)	(0.08)
3	(2) + weather	10	0.0081	4.24	582.4
			(0.002)	(0.12)	(0.09)
4	(3) + season	13	0.0077	26.74	555.6
			(0.0004)	(<0.0001)	(0.28)
5	(4) + year/day intervals	20	0.0059	18.54	537.1
			(0.025)	(0.01)	(0.38)
6	(5) + year/season	34	0.0046	16.20	520.9
			(0.08)	(0.32)	(0.41)

[a] COH, coefficient of haze. The coefficient in the table is 19.85 [exp. $(\hat{\beta}_1) - 1$], where $\hat{\beta}_1$ is the estimated COH coefficient from the Poisson regression.

[b] Comparison with the model just above it. For example, −2 log likelihood ratio between models 5 and 4 is 18.54.

[c] Comparison with the saturated model [$E(Y_{it}) = \mu_{it}$].

[d] p-Values are in parentheses.

of the effect is similar to, if not larger than, that found in the London studies (see calculation below).

Same-day COH appeared as a significant predictor except in models 6 and 7. However, in both these models some of the correlation between COH and mortality was filtered out. Model 6 did not add significantly to the fit of model 5. If COH is eliminated from models 5 and 6 (as was done in models 5a and 6a in Table 3), the inclusion of higher-order, year-specific seasonal curves still does not improve model 5a significantly, so it can reasonably be argued that the added variables just added noise. Several stepwise regressions were done, and COH was included in every case. When regressions were done by disease category, respiratory deaths were highly significant, as might be expected if particulates (as opposed to a surrogate) are having an effect on mortality.

It is important to point out that association does not imply causation; any inference about cause and effect must be based on nonstatistical arguments. What can be concluded is that the higher death rate on days with high particulate must have some cause, and that this cause is correlated with high particulate levels. However, it is conceivable that this causal factor would be unaffected by a decrease in particulate levels. For example, days with high particulates might be associated with cold days where people are more likely to stay indoors. Staying indoors might be the real cause of higher mortality. As mentioned in the "Introduction," there is widespread agreement that high particulate levels can be harmful, so it is reasonable to expect this to be at least a contributing cause for Santa Clara County.

To conclude that particulates do actually affect mortality, one is led to ask how COH, which is a measure of particulate levels outdoors, could be affecting those who are dying indoors. Hospitals are generally designed with air filtration systems, but nursing homes and individual homes are generally not tightly sealed to the outside, at least not in Santa Clara County with its mild climate. Several studies

have found a correlation between indoor and outdoor particulate levels. Dockery and Spengler (*11*) found a 70% in filtration rate of respirable particles in buildings not tightly sealed, that is, respirable particulates in such buildings included a component from the outside air, and this component ran about 70% of outdoor levels. Thus, a large segment of people dying are exposed to outdoor particulates.

By showing that there is the possibility that particulates may affect mortality, this study demonstrates that the possibility of a health risk exists at levels found in Santa Clara County. Previous particulate studies, such as Mazumdar et al. (*1*), which this study attempts to parallel, have provided key evidence for setting standards [EPA Criteria document (*12*)]. Although the evidence of Mazumdar et al. was indirect, they raised the reasonable possibility of an effect. This possibility of a health risk at the particulate concentrations found in London between 1958 and 1972 was an important factor in setting the Federal particulate standard. This study suggests that particulates may be a health hazard at much lower levels, even below the national standard. It would be worthwhile to perform similar studies in other areas to see if this effect is widespread. Comparisons among such studies might help determine whether the particulates themselves or some surrogate is actually responsible for the deaths.

It is very difficult to estimate the magnitude of the assumed effect of particulate and to predict the effect of measures to lower particulate levels. The principal obstacles include: *a*) the fact that COH is a surrogate for the real culprit; *b*) the relationship between the concentrations at the monitoring site in San Jose at the 4th Street monitoring site in San Jose and the actual dosage people are getting; *c*) uncertainties with extrapolating the mortality/COH relationship to other seasons and lower concentrations; *d*) the fact that particulate control measures will affect some particulate constituents more than others; and *e*) the fact that the control measures themselves affect people's behavior and hence might directly or indirectly affect people's health. For example, reducing auto traffic would save lives because of fewer traffic accidents.

Unlike most regulated substances, such as lead, NO_2, SO_2, radon, benzene, saccharine, etc., each of which is a specific compound, particulates are a mixture of many substances from many different sources. It is undoubtedly true that some particulates have a more serious health impact than others. A recent study by Thurston et al. (*13*) suggests that acid aerosols are of greater health significance than particulates as a whole. The following argument may lend some indirect support to this.

Several authors (*2,3*) have noticed an odd fact about the dose-response relationship for London. In later years, when particulate levels were lower, the regression slopes are, if anything, steeper. Upon closer inspection, it appears that this effect occurs within individual years, i.e., that the dose-response curve does not become less steep at lower particulate levels. This result is surprising because London's particulate levels had been reduced by a factor of 10, so that it might be expected that these reduced concentrations would reach the lower part of an S-shaped does-response curve.

Santa Clara County has lower particulate levels than London did even with its reductions in the 1960s, and hence the slope for Santa Clara County should represent a still lower part of the dose-response curve. The comparison of coefficients in this study with the London coefficients is rough to say the least, but may still shed some light on the nature of the dose-response curve.

There were about 280 deaths per day in London versus 20 for Santa Clara, The rationale for using the ratio of daily deaths instead of the population ratio is the assumption that the susceptible population may be people who are very weak to begin with. Conversion from British smoke (BS) units to COH is extremely rough, but the California Air Resources Board (*14*) suggests a BS/total suspended particulate (TSP) ratio of 0.55. [This will vary from location to location, and may be on the high side based on Pashel and Egner (*15*), Clayton and Wallin (*16*), and observations from Australia where co-located instruments are run.] The ratio of COH to TSP is at least 1.0 at 4th Street for the years in question. Therefore, a 0.55 BS/COH ratio is, if anything, on the high side.

Finally, it is important to mention that particulate concentrations at the 4th Street site run about one-third higher than most of the other sites in Santa Clara County (although the correlations are generally quite high) based on comparisons among the many Santa Clara County sites which the district operated until 1980. In contrast, the London studies were based on an average of samplers from around the area that are assumed to be representative of the ambient concentrations. Combining these facts we get a conversion factor of $(280/20 \times (1/0.55) \times (4/3) = 33.9$ for translating the Santa Clara/COH coefficient to a London BS coefficient. Thus, the COH regression coefficient from model 4, Table 3, the predicted increase of 0.0084 daily deaths in Santa Clara/COH unit, translates into an increase of $33.9 \times 0.0084 = 0.285$ daily London deaths/$\mu g/m^3$ BS. The model in Ostro (*2*) is somewhat closer to model 4 than Schwartz and Marcus (*3*). Ostro computed regression slopes for each year in London and for BS concentrations below 150 $\mu g/m^3$. These slopes had no apparent trend, and their median was 0.128 daily deaths/BS unit. Thus, the Santa Clara slope appears steeper than the slope for London.

There is great uncertainty in the conversion factor of TSP to BS. However, no published study has found BS/TSP ratios larger than 1. Using an upper bound of 1 instead of 0.55, the Santa Clara coefficient is still larger than the slope for London. If the population ratio is used instead of daily deaths (and the 0.55 factor is used), the translated Santa Clara coefficient is quite similar to London's. Hence, even given the extreme roughness of the conversion factor, it appears that the Santa Clara County coefficient is at least as large as London's.

These anomalous findings raise the distinct possibility that the variable selected for this study, COH, is acting as a surrogate for some constituent particle or particles that are the real cause of the health effects. One consistent hypothesis is that this constituent increases logarithmically with total particulates in London and is a larger proportion of the particulate mix in Santa Clara County than in London.

This possibility has regulatory consequences. For example (for argument's sake only) suppose that nitrates were the real culprit. Then regulations which reduced auto traffic or which mandated cleaner-burning engines would be beneficial, while regulations restricting construction dust would not. This example highlights a possible danger imposing particulate regulations on total particulates.

Having stated the caveats, it is also important to stress that the finding of any health effect indicates that there may be a sizeable health risk from particulates. If the mortality signal were not strong enough, no statistically significant effect would be observed. By way of comparison, although temperature certainly affects mortality, the apparent effect of temperature was not statistically significant in many of the models fit. While it is misleading to try to quantify the risk exactly, it can be said with confidence that the effect is on the order of at least one in a thousand early deaths due to particulate concentrations in excess of the State standard. For the San Francisco Bay Area, this risk is much greater than the risks from any toxic identified so far.

This study suggests that particulates may be a health risk at concentration slower than previously suspected. Further study is indicated to attempt to identify which constituents (if any) within the particle mix are causing the health problems. Speciated particulate data could be collected and compared with mortality in a multiple regression. Perhaps other measures such as emergency room visits for respiratory problems or absences from school might be a more sensitive dependent variable than mortality. Similar studies in other areas would help to clarify whether or not particulates are the culprit.

The opinions expressed here are those of the author and not necessarily those of the board of Directors or staff of the Bay Area Air Quality Management District.

The author thanks Alice Whittemore, Bart Ostro, and Herb Schimmel for their helpful comments and criticism. Thanks also to Louise Detwiler of the Santa Clara Department of Public Health for providing the mortality data.

Appendix

Poisson Regressions

The most general model fit was

$$\begin{aligned} E(Y_{it}|\text{predictors}) &= \exp[\ \beta_0 + \beta_1 C_{it} + \beta_2 T_{it} + \beta_3 H_{it} \\ &\quad + \gamma_{i0} + \gamma_{i1}(t\text{-}46.5) + \gamma_{i2}(t\text{-}46.5)^2 \\ &\quad + \gamma_{i3}(t\text{-}46.5)^3\ ] \\ &\equiv \exp[\ \boldsymbol{\beta}'\mathbf{x}_{it}\ ], \end{aligned}$$

where Y_{it} is the mortality on day t in year i, and the Y_{it} values are a sequence of independent random variables.

$$\boldsymbol{\beta}' = (\beta_0,\beta_1,\beta_2,\beta_3,\gamma_{10},\gamma_{20},\ldots,\gamma_{83}), \text{ and}$$

$$\mathbf{x}_{it} = (1,C_{it},T_{it},H_{it},I_{it}(1),I_{it}(1)(t\text{-}46.5),\ldots,I_{it}(8)(t\text{-}46.5)^3),$$

and where the C_{it}, T_{it}, etc., are as before. The model was fit by maximizing the likelihood function

$$L(\boldsymbol{\beta}) = \prod \exp[\ y_{it}\boldsymbol{\beta}'\mathbf{x}_{it} - \exp(\boldsymbol{\beta}'\mathbf{x}_{it})\ ]/y_{it}!$$

with maximum likelihood estimates $\hat{\boldsymbol{\beta}}$. Models can be compared on the basis of increase in the likelihood. Specifically, let β_0 be the vector of parameters from submodel with p_0 parameters, contained within another model with p_1 parameters β_1. Then the likelihood ratio test statistic

$$-2\ln[\ L(\hat{\boldsymbol{\beta}}_0)\ /\ L(\hat{\boldsymbol{\beta}}_1)\ ]$$

is approximately χ^2 with $p_1 - p_0$ degrees of freedom. Table 5 in the text contains these χ^2 values comparing each model with the largest submodel above it.

To make β_1 comparable to the coefficients in Tables 3 and 4, where the regression coefficient represents the Δ = change in expected daily deaths per change in 1 unit COH, let $\mathbf{x}_{it}^*$ be $\mathbf{x}_{it}$ only with C_{it} replaced with $C_{it} + 1$. Then

$$\begin{aligned} \Delta &= \exp(\boldsymbol{\beta}'\mathbf{x}_{it}^*) - \exp(\boldsymbol{\beta}'\mathbf{x}_{it}) \\ &= \exp(\boldsymbol{\beta}'\mathbf{x}_{it})[\exp(\hat{\beta}_1) - 1] \\ &= E(Y_{it}|\mathbf{x}_{it})[\exp(\hat{\beta}_1) - 1]. \end{aligned}$$

While $E(Y_{it}|\mathbf{x}_{it})$ varies somewhat with $\mathbf{x}_{it}$, it is generally near the daily mortality rate of 19.85. Thus

$$19.85[\exp(\hat{\beta}_1) - 1]$$

is roughly comparable to previous COH coefficients. These values are presented in Table 7.

For the significance level of the COH coefficient, I fit the model without COH and compared the models with and without using likelihood ratio test. The significance level is the *p*-value of this test.

In addition to taking the underlying distribution of the NOAC series into account, the Poisson regression has the advantage of providing a test of model adequacy. Any Poisson regression model can be compared to the saturated model where $E(Y_{it}) = \mu_{it}$. Here the maximum likelihood estimate is $\hat{\mu}_{it} = Y_{it}$ and the likelihood ratio statistic

$$-2\ln[\ L(\hat{\boldsymbol{\beta}})/L(\hat{\boldsymbol{\mu}})\ ]$$

is approximately χ^2 with n−p degrees of freedom. The last column of Table 7 gives the corresponding x^2 and *p*-values.

The results of Table 7 are quite similar to the results in Table 3. The COH coefficient is not strongly affected by temperature, humidity, year effect, or seasonal terms, but decreases when the year-specific trend or season terms are added. It is noteworthy that models 4 and higher fit adequately, that is, the remaining variation in the Y_{it} could easily be due to chance.

REFERENCES

1. Mazumdar, S., Schimmel, H., and Higgins I. T. T. Relationship of daily mortality to air pollution: an analysis of 14 London winters 1985/59-1971/72. Arch. Environ. Health 37 (4): 213–220 (1982).
2. Ostro, B. D. A search for a threshold in the relationship of air pollution to mortality: a reanalysis of London winters. Environ. Health Perspect. 58: 397–399 (1984).
3. Schwartz, J., and Marcus, A. Statistical Reanalysis of Data Relating Mortality to Air Pollution during London Winters 1985–1972. Working Paper, U.S. Environmental Protection Agency, Washington, DC, October, 1986.
4. Glasser, M., and Greenberg, L. Air pollution and mortality and weather, New York City, 1960-64. Arch. Environ. Health 22: 334–543 (1971).
5. Schimmel, H., and Greenberg, L. A study of the relationship of pollution to mortality, New York City, 1963–1968. J. Air Pollut. Control Assoc. 22: 607–616 (1972).
6. Schimmel, H., and Murawski, T. J. The relation of air pollution to mortality. Occup. Med. 18: 316–333 (1976).
7. Chestnut, L. G., Rowe, R. D., and Ostro, B. D. Santa Clara Criteria Air Pollutant Benefit Analysis. Prepared for the U.S. Environmental Protection Agency under Contract No. 68-01-7033, Work Assignment No. 216, Energy and Resource Consultants, Boulder, CO, 1987.
8. U.S. Bureau of the Census. Statistical Abstract of the United States 1988, 108th ed. Washington, DC, 1987.
9. Bay Area Air Quality Management District Base Year 1983 Emissions Inventory Summary Report. Bay Area Air Quality Management District, San Francisco, CA, 1987.
10. California Air Resources Board, Identification of Particulate Matter Species Profiles, Vol 2. California ARB, Sacramento, CA, 1989.
11. Dockery, D.W., and Spengler, J. D. Indoor-outdoor relationships of respirable sulfates and particles. Atmos. Environ. 15: 333–343 (1981).
12. U.S. EPA. Air Quality Criteria for Particulate Matter and Sulfur Oxides III. Office of Research and Development, Environmental Criteria and Assessment Office, Research Triangle Park, NC, 1982.
13. Thurson, G. D., Ito, K., Lippmann, M., and Hayes, C. Reexamination of London, England, mortality in relation to exposure to acidic aerosols during 1963-1972 winters. Environ. Health Perspect. 79: 73–82 (1989).
14. California Air Resources Board. California Ambient Air Quality Standard for Particulate Matter (PM10). California ARB, Sacramento, CA, December 1982.
15. Pashel, G. E., and Egner, D. R. A comparison of ambient suspended particulate matter concentration as measured by the British smoke shade sampler and the high volume sampler at 16 sites in the United States. Atmos. Environ. 15: 919–927 (1981).
16. Clayton, P., and Wallin, S. C. Comparison of ambient suspended particulate matter concentrations as measured by the British smoke sampler and the high volume sampler. Atoms. Environ. 16: 2031–2032 (1982).