primarily sulfate particles in the eastern United States, and primarily nitrates in the western United States. In the eastern U.S., the sulfate particles sometimes occur in the form of sulfuric and other acids. These secondary particles are usually between 0.3 and 1 μm in aerodynamic diameter, and so differ both in size distribution and in chemical composition from the primary combustion particles. Because these different classes of particles have different sources, different regulatory policies are required to reduce their concentrations. If the reported associations between particulate air pollution and daily mortality are due principally to one class, this has obvious public policy implications.

The distribution and deposition of particles in the lung varies substantially with particle size. Coarse particles have a higher probability of being deposited in the bronchial region.[7] Fine particles have a higher probability of being deposited in the periphery of the lung, especially in the respiratory bronchioles and alveoli, where their clearance is slow relative to particles deposited on airways.[10,11] The deposition of particles is enhanced in patients with chronic obstructive pulmonary disease (COPD)[12,13] and in healthy women.[14] The deposition of particles does not directly vary with age, but does vary with breathing rate and pulmonary function level.[15] Hence, identification of the size range responsible for the observed associations would also provide insight into possible mechanisms of action.

Fine and coarse particle mass and sulfate concentrations were monitored in six eastern U.S. cities over an eight-year period. A previous report[16] analyzed the association of daily mortality with various measures of particulate air pollution in two of these cities (St. Louis, MO and Harriman, TN) for the one-year period when aerosol acidity measures were available. In this analysis, we sought to confirm the association between daily variations in airborne particles measured as PM_{10} and daily variations in deaths in four additional locations, and in additional years for the two previously analyzed locations. The primary goal, however, was to determine whether the associations between particulate air pollution and daily mortality were specifically due to fine particles. Third, we sought to determine whether the associations with fine particle mass were primarily attributable to the sulfate or acidity of these particles. Finally we sought to determine whether risks were elevated for certain causes of death or for the elderly.

DATA AND METHODS

The Harvard Six Cities Study was initiated in 1974 as a prospective study of respiratory symptoms and pulmonary function in population-based cohorts of adults and elementary school children in each of six U.S. communities: Watertown, MA; Kingston and Harriman, TN; St. Louis, MO; Steubenville, OH; Portage, WI; and Topeka, KS.[17] As part of that study, central residential ambient air pollution monitoring stations were established in each community.

Exposure Assessment

Beginning in 1979, dichotomous virtual impactor samplers were placed at the central residential monitoring sites in each of the six cities.[18] Separate samples were collected of fine particles ($PM_{2.5}$) and of the coarse mass (CM) fraction (aerodynamic diameter greater than 2.5 μm and less than 15 μm before 1984, and less than 10 μm after 1984). The sum of the fine and coarse mass is the inhalable particulate mass (PM_{15} or PM_{10}). The concentrations of PM_{15} and PM_{10} were very similar within each city before and after the change in the upper size cut, and we will therefore treat these as PM_{10} measurements. Integrated 24-hour samples were collected daily for part of the study periods, but were collected at least every other day until the late 1980s (Table 1). Mass concentration was determined by beta-attenuation, and elemental composition was determined by x-ray fluorescence; these measurements were made separately for the fine and coarse particle samples.[19] Elements routinely measured included sulfur, which allowed the estimation of sulfate ($SO_4^=$) concentrations from the fine mass fraction. For a two-year period, 1983 to 1984, mass concentrations were determined gravimetrically, and sulfate concentrations were determined turbidimetrically.[20] During a shorter period (Table 1), the hydrogen ion concentration of the fine particle mass was determined using the Harvard denuder system.[21] The separate measurements of fine and coarse mass allow us to assess whether the previously reported associations between airborne particles and daily deaths are primarily due to particles of one size range.

Table 1. Start dates, stop dates, and number of samples by measured particle size and city, Harvard Six Cities Study, 1976 to 1987.

Study	*Sampling*	*$PM_{2.5}$ and $PM_{10/15}$*			*H^+*		
Community	*Site*	*Start*	*End*	*Samples*	*Start*	*End*	*Samples*
Boston, MA	Watertown, MA	2 May 79	2 Jan 86	1140	20 Jun 87	27 Aug 88	155
Knoxville, TN	Harriman, TN	30 Jan 80	31 Dec 87	1481	6 Dec 85	21 Aug 86	232
St. Louis, MO	Carondelet, MO	22 Sep 79	19 Jan 87	1375	15 Oct 86	28 Dec 87	429
Steubenville, OH	Steubenville, OH	13 Apr 79	26 Sep 87	1520	6 Dec 85	4 Sep 86	223
Madison, WI	Portage, WI	22 Mar 79	31 Dec 87	1436	31 Oct 86	13 Sep 87	301
Topeka, KS	Topeka, KS	23 Sep 79	18 Oct 88	1432	9 Nov 87	30 Aug 88	281

Meteorological Data

Meteorological data was obtained from the National Center for Atmospheric Research. Hourly measures of temperature, dewpoint temperature, and precipitation were obtained from the National Oceanographic and Atmospheric Administration weather station closest to each city (Table 2). The hourly measures for each city were collapsed over 24-hour periods to provide a maximum, minimum, and mean value for temperature and dewpoint temperature, and to give indicator variables for rain or snow.

Mortality Data

The base population for each of the six metropolitan areas was defined as the county in which the air pollution monitor was located plus adjacent contiguous counties. The counties defining the metropolitan areas for each location were defined as shown in Table 2. Each metropolitan area was identified by the name of the largest city included in the study area.

Daily deaths were extracted from annual detail mortality tapes from the National Center for Health Statistics for the time periods when fine and coarse particulate samples were collected. Extracted information for deaths included date and county, age, sex, and underlying cause (International Classification of Disease, Ninth Revision [ICD9]). Deaths due to accidents and other external causes (ICD9 800-999) were excluded. Separate counts were also computed for deaths of persons age 65 and over and for deaths from ischemic heart disease (ICD9 410-414), pneumonia (ICD9 480-486), and COPD (ICD9 490-496).

Analytic Methods

Counts of daily deaths were modeled by Poisson regression. In a standard Poisson regression, we assume that

$$\log (E(Yi)) = Xi\beta \tag{1}$$

where Y_i is the number of deaths on day i, X_i is the vector of covariates on day i, β is the vector of regression coefficients, and E denotes expected value.

Because the unit of analysis in this study is the day, the potential confounders that must be controlled are those that vary over time, possibly in coincidence with air pollution. There have been long-term trends in mortality rates over time, and mortality is known to vary substantially by season. These trends are primarily due to factors other than air pollution. Since many of those factors are unmeasured (e.g., smoking rates) long-term trends pose a potential problem for this type of study. We have dealt with this potential by removing all long-term time trends from the data. There is no reason to believe that day-to-day fluctuations in factors such as smoking or changes in medical practice are correlated with day-to-day changes in air pollution, so this approach should remove that potential for confounding.

Neither the time trends nor the weather effects need be linear. While the standard Poisson regression analysis can incorporate nonlinear dependencies, it is necessary to specify the functional form of those dependencies. If we believe we know the correct covariates but not necessarily the correct functional forms, the generalized additive model offers a nonparametric regression alternative. It assumes

$$\log (E(Y)) = \sum Si\ (Xi) \tag{2}$$

In this case $S_i(X_i)$ is a smooth function of X_i that is fit nonparametrically. A nonparametric smoother is a tool for summarizing the trend of a response measurement Y as a function of one or more predictor measurements. The properties of such smoothers have been extensively discussed.[22,23] The smoothers are generalizations of a weighted moving average. In a weighted moving average, a window or neighborhood is chosen about each point X_i. Generally, it is chosen to include a fixed percentage of the observations that are closest to X_i. Then the weighted mean of the outcome measure in the window is computed. This is the smoothed estimate of outcome at point X_i, and the method is repeated for all observations. A variant which performs better at the extremes of the data is loess,[22] which performs a weighted

Table 2. Counties included in analysis and weather station used for data extraction.

	Boston, MA	*Knoxville, TN*	*St. Louis, MO*	*Steubenville, OH*	*Portage, WI*	*Topeka, KS*
Counties Included	Suffolk Middlesex Norfolk	Anderson Cumberland Knox Loudon Meigs Morgan Roane Rhea	Clinton, IL Madison, IL Monroe, IL St. Clair, IL Franklin, MO Jefferson, MO St. Charles, MO St. Louis, MO St. Louis, City	Brooke, WV Hancock, WV Jefferson, OH	Adams Columbia Dane Dodge Green Lake Jefferson Juneau Marquette Sauk	Shawnee
1980 Population	2,274,202	640,887	2,356,460	163,734	324,990	154,916
NOAA Weather Station	Logan Airport, Boston	Knoxville Airport	Lambert Field, St. Louis	Greater Pittsburgh Airport	Madison	Municipal Airport, Topeka

Table 3. Percentile points of the pollution and weather variables included in the analysis.

Pollutant	*# Obs*	*5%*	*25%*	*50%*	*75%*	*95%*
$PM_{2.5}$ (µg/m³)	8432	4.3	9.0	14.7	23.0	431
CM (µg/m³)	8387	1.0	5.0	9.0	15.5	30.1
PM_{10} (µg/m³)	8374	8.0	16.0	25.0	38.0	67.8
$SO_4^=$ (µg/m³)	8409	1.5	3.3	5.8	10.2	22.3
H^+ (nM/m³)	1621	0	4.1	10.2	23	67.3
Temp (°C)	18620	-7.5	3.2	12.4	20.5	26.4
Dewpoint (°C)	18605	-13.4	-2.5	6.2	14.7	20.7

regression in each moving window. The weights decline to zero at the boundaries of the window, and are near one in the central third of the window. Such approaches have been used in many recent studies relating air pollution with mortality[2,23,24] and with hospital admissions.[3,5] Generalized additive models were used to estimate city-specific Poisson regressions of daily mortality with time trends, weather variables, and particulate air pollution.

In nonparametric regression, a smoothing parameter must be chosen which is the percentage of the data that will be included in each neighborhood. For the smooth against time, we have chosen to use 5% of the data in each neighborhood, to ensure removal of all long wavelength patterns from the data. This gives a neighborhood of about 100 days, with most of the weight concentrated in the central 30 days of the neighborhood. In the hydrogen ion analysis, for which we had fewer days of data, we chose a parameter to give us a neighborhood with about the same number of days as the other analyses. We tested the adequacy of this assumption by examining, in each location, spectral density functions for the residuals of regression models without air pollution. For the smooth functions of temperature and dewpoint temperature, we chose to use 50% of the data, because we expected the dose-response relationship between those variables and mortality to be much smoother. Analyses were performed to determine the sensitivity of the air pollution effect to weather modeling.

Our interest is in a global estimate of the association between different measures of airborne particles and daily mortality. However, the temporal and weather dependence of mortality may vary among the six metropolitan areas. Therefore, separate regression models were fit in each area, and combined regression coefficients were calculated in a second-stage analysis using inverse variance weighting. In each location, we regressed the daily number of deaths on smooth functions of day of study, temperature, dewpoint temperature, indicator variables for rain and snow, and day-of-week indicators. All regressions were estimated in S-PLUS,[25] using the generalized additive model function.

Previous analyses have shown the strongest associations of daily mortality with particle concentrations on the same day and on the previous day. In these data, observations generally were measured only every other day, except during special intensive monitoring periods. For $PM_{2.5}$, 62% of the samples did not have observations on the previous day. Therefore it was not possible to investigate lag structures in these data in any detail. Based on previous observations, mortality was assumed to be associated with the mean of the non-missing particle concentrations on the same and on the previous day. This had the effect of increasing the number of days included from 8,432 days with $PM_{2.5}$ measurements to 12,055 days with $PM_{2.5}$ measurements on the same or on the previous day.

Relative risks have been expressed as the percent increase in daily mortality (relative risk minus one times 100) associated with each specified increment in particulate air pollution exposure. A common choice of increment has been 10 mg/m³, and we have used that increment in our reporting. However, some of our analyses compare the results using several different measures of airborne particles, some of which are nested within others. For example, fine sulfate mass is a subset of all fine mass. We wish to determine which exposure measure is a better marker for the relevant particle exposure. This information is not conveyed in the slope of effect per 10 µg/m³ of exposure, since the range of exposure differs for the different particle measures. If the change in daily deaths associated with a change in exposure across the range of exposure for fine sulfate mass, for example, is as large or larger than for the range of daily deaths associated with the range of exposure to $PM_{2.5}$, it would suggest that the association was specifically with the sulfates, or with a subset of them. To avoid undue influence by a few outliers in exposure, we have examined the percent increase in daily deaths associated with an increase in the 5th to 95th percentile of exposure for each particle measure. In addition to total mortality from all causes, we examined mortality of persons age 65 and older to determine whether age modifies the effect of air pollution, and causes specific mortality for ischemic heart disease, COPD, and pneumonia.

RESULTS

The overall distribution of the air pollution, weather, and mortality data in the study is shown in Table 3. PM_{10} concentrations were generally well below the 24-hour ambient air quality standard of 150 µg/m³. The means and standard deviations of air pollution, weather, and mortality data in the six metropolitan areas individually are shown in Table 4. In all cities except Topeka, the majority of PM_{10} mass was in the fine particle mass. $PM_{2.5}$ concentrations were approximately 65% of the PM_{10} mass concentrations in four cities (Table 4), 62% in St. Louis, and only notably lower in Topeka (50%). The cities also differed in the correlation between the two different size ranges of particles (Table 4). In Boston, Portage, and Topeka the correlation between $PM_{2.5}$ and CM was low (~0.3), while in Steubenville, it was quite high (0.7).