ity in 10 cities across the United States.

Data

To effectively analyze the distributed lag between particulate matter <10 μm/100 μm in diameter (PM_{10}) and daily deaths, we obtained daily PM_{10} measurements. Most of the US cities measured PM_{10} only on 1 day of 6, but a number of locations monitored daily. We selected 10 US cities with daily PM_{10} monitoring to provide a reasonable number of locations for a combined analysis: Birmingham, Alabama; Canton, Ohio; Chicago, Illinois; Colorado Springs, Colorado; Detroit, Michigan; Minneapolis–St. Paul, Minnesota; New Haven, Connecticut; Pittsburgh, Pennsylvania; and Seattle and Spokane, Washington.

Daily counts of deaths due to pneumonia (International Classification of Diseases, 9th Revision [ICD-9]: 480–487), chronic obstructive pulmonary diseases (COPD) (ICD-9: 490–496), all cardiovascular diseases (ICD-9: 390–429), and specifically myocardial infarction (ICD-9: 410) in the metropolitan county containing each city were extracted from National Center for Health Statistics mortality tapes for the years 1986 through 1993. Minneapolis and St. Paul were combined and treated as one city. Daily weather data were obtained from the nearest airport station (EarthInfo CD NCDC Surface Airways; EarthInfo Inc, Boulder, CO), and daily concentrations of PM_{10} were obtained from the US Environmental Protection Agency's Aerometric Retrieval System monitoring network.

Many of the cities have more than one monitoring location for PM_{10}. Some operate on a daily basis; others on every third or sixth day. If the monitoring results were simply averaged, the daily mean would change on days when new monitors were included merely because their annual average differed from the monitoring station that operateed on a daily basis. The variance of PM_{10} measurements also can differ from monitoring location to location. Day-to-day changes in which monitoring results are averaged would also result in changes in the day-to-day variation in the exposure measurements that would not represent true changes in exposure but rather changes in the sampling of the monitors. To remove these influences, we used the following algorithm. The annual mean was computed for each monitor for each year and was subtracted from the daily values of that monitor. We then standardized these daily deviances from each monitor's annual average by dividing by the standard deviation of that monitor. The daily-standardized deviations for each monitor on each day were averaged, producing a daily averaged standardized deviation. We multiplied this by the standard deviation of all centered monitor readings for the entire year, and we added back in the annual average of all of the monitors. This gave a daily average PM_{10} concentration for each day in each city. This approach has been applied successfully in previous analyses.[9,11]

Analytical Methods

For each city, a generalized additive Poisson regression was fit,[12,13] modeling the logarithm of the expected value of daily deaths as a sum of the smooth functions of the predictor variables. The generalized additive model allows regressions to include nonparametric smooth functions to model the potential nonlinear dependence of the daily admissions on weather and season. It was assumed that:

$$\log(E(Y)) = \beta_0 + S_1(X_1) + \ldots + S_p(X_p), \quad (1)$$

where Y is the daily count of deaths, $E(Y)$ is the expected value of that count, X_i is the covariate, and S_i is the smooth (ie, continuously differentiable) function. For the S_i we used loess,[14] a moving regression smoother. This approach is now standard in air pollution time-series.[15] For each covariate, it is necessary to choose a smoothing parameter that determines how smooth the function of that covariate should be. Three classes of predictor variables were used: a smooth function of time to capture seasonal and other long-term trends in the data, weather and day of the week variables to capture shorter-term potential confounding, and PM_{10}. The choice of smoothing parameter for each set of variables is described below. The purpose of the smooth function of time is to remove the basic long-term pattern from the data. Seasonal patterns can vary greatly between Birmingham and Spokane, for example, and a separate smoothing parameter was chosen in each city to eliminate seasonal patterns in the residuals and to reduce the residuals of the regression to "white noise"[16] (ie, remove serial correlation). This approach was used because each death is an independent event, and autocorrelation in residuals indicates there are omitted, time-dependent covariates whose variation may confound air pollution. If the autocorrelation is removed, the remaining variation in omitted covariates has no systematic temporal pattern; hence, confounding is less likely. This approach was described previously.[16] Sometimes it was necessary to incorporate autoregressive terms[17] to eliminate serial correlation from the residuals.

The other covariates were temperature, relative humidity, and barometric pressure on the same day; the previous day's temperature; and the day of the week. To allow for city-specific differences, the smoothing parameters for these covariates were also chosen separately in each location to minimize Akaike's Information Criteria.[18]

PM_{10} was treated as having a linear association with deaths in this analysis to facilitate the combination of coefficients across cities and the examination of lag structure. Robust regression was used to reduce sensitivity to outliers in the dependent variable. To reduce sensitivity to outliers in the pollution variable, the

TABLE 1
The Populations and Daily Means of Deaths in the 10 Locations

Cities	1990 Population	Deaths*				
		Total	Pneumonia	COPD	CVD	MI
Birmingham	907,810	19.1	0.6	0.7	5.9	1.5
Canton	367,585	9.9	0.3	0.5	3.5	1.2
Chicago	5,105,067	133.4	5.0	4.1	48.4	15.2
Colorado Springs	397,014	6.0	0.3	0.4	1.6	0.5
Detroit	2,111,687	59.7	1.8	1.9	22.8	6.2
Minneapolis	1,518,196	32.3	1.3	1.4	9.7	2.4
New Haven	804,219	20.4	0.8	0.7	7.5	1.7
Pittsburgh	1,336,449	42.4	1.3	1.6	16.1	5.1
Seattle	1,507,319	29.3	1.1	1.4	9.4	2.3
Spokane	361,364	8.7	0.3	0.5	2.9	0.9

* COPD, chronic obstructive pulmonary disease; CVD, all cardiovascular diseases; MI, myocardial infarction.

baseline analysis was restricted to days when PM_{10} levels were below 150 $\mu g/m^3$, the currently enforced ambient standard. This also ensured that the results were unambiguously relevant to questions of revision of those standards.

Distributed Lag Models

Distributed lag models have been used for decades in the social sciences,[19] and Pope and Schwartz[20] recently described the use of this approach in epidemiology. We recently applied this method to estimating the distributed lag between all-cause mortality and daily deaths in these 10 locations.[9] The motivation for the distributed lag model is the realization that air pollution can affect deaths occurring not only on the same day but also on several subsequent days. The converse is therefore also true: deaths today will depend on the same-day effect of today's pollution levels, the 1-day-lag effects of yesterday's PM_{10} concentrations, and so forth. Therefore, suppressing covariates and assuming Gaussian data in this instance, the *unconstrained* distributed lag model assumes:

$$Y_t = \alpha + \beta_0 X_t + \ldots + \beta_q X_{t-q} + \varepsilon_t, \quad (2)$$

where X_{t-q} is the PM_{10} concentration q days before the deaths. The overall effect of a unit increase in air pollution on a single day is its impact on that day plus that on subsequent days; that is, it is the sum of $\beta_0 + \ldots + \beta_q$.[16] To simplify this, equation 2 can be recast as:

$$Y_t = \alpha + \beta^*(\omega_0 X_t + \ldots + \omega_q X_{t-q}) + \varepsilon_t, \quad (3)$$

where the ω_i are weights that sum to 1, and β^* is $\beta_0 + \ldots + \beta_q$. That is, β^* is also interpretable as the marginal effect of a unit increase in a weighted average pollution variable. Because a unit of increase in pollution on a single day increases the weighted average on all q subsequent days, the effect of that single day's increase will be $\beta^*\omega_i$ on each of the q subsequent days, or β^* overall.

Because there is substantial correlation between air pollution concentrations on days that are close together, the above regression will have a high degree of collinearity. This will result in unstable estimates of the individual β_q's and, hence, poor estimates of the shape of the distribution of the effect over lag. However, the sum of the individual β_q's are an unbiased estimate of the overall effect of a unit increase in pollution, and the individual β_q's are also unbiased. Thus, by fitting the same model in 10 different locations, and by combining effect size estimates, by lag, over the cities, we can obtain an unbiased estimate of the distribution of the effect over time with no constrains. Note that the use of a single day's exposure is a constrained lag model. In that case, we fit equation 2 with the constrain that $\beta_1 = \beta_2 = \ldots = \beta_p = 0$. If we are unsure that the pollution effects are limited to a single day, these constrains are quite restrictive and are therefore likely to introduce bias into the estimated overall effect. To see if the traditional approaches of using a 1- or 2-day moving average resulted in a downward bias in the estimated effects of PM_{10}, we also fit those models in each city.

To combine results across cities, we used inverse variance-weighted averaging to estimate the overall effect for lag 0, lag 0–1. For the distributed lag model, we used a multivariate approach to pool the estimates of each day's lag simultaneously. The inverse variance-weighted average (and overall covariance matrix) was computed using the variance–covariance matrix of the estimated coefficients in each city. This approach incorporated the correlations among the estimated effects at different lags.

Results

Table 1 shows the populations and the mean values of the cause-specific deaths in the 10 study locations. Table 2 shows the mean values of the environmental variables in these cities.

As presented previously,[9] the correlation between PM_{10} and barometric pressure was quite small, ranging