The Concentration-Response Relation
Between PM$_{2.5}$ and Daily Deaths.

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th>Citation</th>
<th>Schwartz, Joel, Francine Laden, and Antonella Zanobetti. 2002. The concentration-response relation between PM$_{2.5}$ and daily deaths. Environmental Health Perspectives 110(10): 1025-1029.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td><a href="http://ehp03.niehs.nih.gov/home.action">http://ehp03.niehs.nih.gov/home.action</a></td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:8000921">http://nrs.harvard.edu/urn-3:HUL.InstRepos:8000921</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
The Concentration–Response Relation between PM$_{2.5}$ and Daily Deaths

Joel Schwartz, Francine Laden, and Antonella Zanobetti

Abstract

In the last decade, a series of studies reported associations between daily concentrations of airborne particles and daily deaths (1–3). The magnitude of the regression coefficients in those studies indicated that particulate air pollution was associated with between 50 and 100,000 early deaths per year in the United States, and similar numbers were found in Europe. More recently, a number of large, multicity studies (4–7) have reported associations between airborne particles, measured in various ways, and daily deaths. The largest study demonstrated that gaseous air pollutants did not confound the association, and that none of the gaseous air pollutants showed an independent effect on daily deaths (7). These studies assumed a linear concentration–response relation between airborne particles and daily deaths and did not address the question of what the association looked like for particle constituents, characterized by size, physiochemical composition, or source.

In a recent study of six U.S. cities (5), we demonstrated that daily mortality was associated with fine particulate matter (aerodynamic diameter $\leq 2.5$ µm; PM$_{2.5}$) and not with coarse particulate matter (aerodynamic diameter between 2.5 and 10 µm; PM$_{2.5–10}$). Each 10 µg/m$^3$ increase in the 2-day mean concentration of PM$_{2.5}$ was associated with a 1.5% (95% confidence interval, 1.1–1.9%) increase in daily mortality.

Ambient PM$_{2.5}$ consists mainly of combustion particles from motor vehicles and the burning of coal, fuel oil, and wood, but also contains some crustal particles from finely pulverized road dust and soils. These sources produce particles with different characteristics, and the relative toxicity of those sources and characteristics is an area of relative recent but intense interest. In a follow-up study (8), we used the elemental composition of size-fractionated particles to identify several distinct source-related fractions of fine particles. We then examined the association of these fractions with daily mortality in each of the six cities and combined the city-specific results in a meta-analysis to derive overall relative risks for each fraction. We found positive associations with particles from traffic, particles from coal, and particles from residual oil combustion when included jointly in the model predicting daily deaths (8). The largest effect size was for residual oil particles, followed by traffic particles and then coal particles. Only the latter two associations were statistically significant, however. Again, as traditional, these analyses assumed a linear association between the various particle constituents and daily deaths.

The shape of the concentration–response relationship is critical for public health assessment, and in particular, some have speculated that thresholds might exist.

Recently, three reports have explored this question for particulate air pollution, using multicity studies in the United States. In one study, Daniels et al. (9) used data from 20 U.S. cities, five of which had daily measurements of PM$_{10}$, with the rest having measurements only one day in six. They used regression splines to model the concentration–response curve in each city and combined the results across cities. They found no evidence for a threshold. In fact, the concentration–response relation was quite linear across the entire range of exposure. In another report, Schwartz and Zanobetti (10) used data from 10 cities, all of which had daily measurements of PM$_{10}$, resulting in slightly more days of study than in the first report. They used nonparametric smoothing to model the concentration–response curve between air pollution and daily deaths in each city and combined the results across cities. Again, a linear, no-threshold relationship was seen. Schwartz and Zanobetti also performed simulations to confirm the ability of this approach to detect thresholds and other types of nonlinearity (11). Schwartz et al. (12), using data from eight Spanish cities, similarly reported a linear association between daily deaths and black smoke, an optical measure of black particles. These results held after adjusting for SO$_2$. To date, no similar examination of the concentration–response curve has been done for PM$_{2.5}$, or for any source components. Because PM$_{2.5}$ is now the regulated form of particulate air pollution in the United States, we here report results of such an analysis.

Materials and Methods

Air pollution data. As part of the Harvard Six Cities studies (12), dichotomous virtual impactor samplers were placed at a central residential monitoring site in six U.S. metropolitan areas: Boston, Massachusetts; Knoxville, Tennessee; St. Louis, Missouri; Stuebenville, Ohio; Madison, Wisconsin; and Topeka, Kansas. Separate filter samples were collected of fine particles (PM$_{2.5}$) and of the coarse mass (PM$_{2.5–10}$) fraction. Integrated 24-hr samples were collected at least every other day from 1979 until the late 1980s, with daily sampling during health survey periods. For fine and coarse particle samples, mass concentration was determined separately by beta-attenuation (13). Except for a period

Address correspondence to J. Schwartz, Environmental Epidemiology Program, Harvard School of Public Health, 665 Huntington Ave., Boston, MA 02115 USA. Telephone: (617) 384-8752. Fax: (617) 384-8745. E-mail: jschwartz@hsph.harvard.edu

This work was supported by U.S. EPA Grant R827353 and NIEHS Grant ES 00002.

Received 21 December 2001; accepted 20 March 2002.
between October 1981 and January 1984 in all cities, elemental composition of fine and coarse mass was determined by X-ray fluorescence (14). Elemental composition was available on 97% of these samples. In the fine fraction, 15 elements were routinely found above the limit of detection: silicon, sulfur, chlorine, potassium, calcium, vanadium, manganese, aluminum, nickel, zinc, selenium, bromine, lead, copper, and iron.

**Source identification.** In separate analyses for each city, we used specific rotation factor analysis to identify up to five common factors from the 15 specified elements. We specified a single element as the tracer for each factor and maximized the projection of these elements using the Procrustes rotation, a variant of the oblique rotation method (15). The Procrustes method allows us to use known tracer for different sources as targets for the different factors and to maximize their loadings on those factors instead of having factors defined in an entirely data-driven manner. To rescale the factor scores from the normalized scale to the mass scale (in micrograms per cubic meter), we regressed the total daily fine particle concentrations on the daily factor scores for all of the factors in separate regression models for each city and took the product of each factor score with its regression coefficient (16). Only sources that were significant predictors of total fine particle mass (p < 0.10) were considered in the mortality analyses. Further details have been published previously (8).

**Meteorologic data.** We obtained meteorologic data from the National Center for Atmospheric Research, including hourly measures of temperature, dew point temperature, and precipitation from the National Oceanographic and Atmospheric Administration weather station nearest to each city (17). We calculated 24-hr mean values for temperature and dew point temperature.

**Mortality data.** We defined the six metropolitan areas in this study as the county containing the air pollution monitor and contiguous counties (5). We extracted daily deaths from annual detail mortality tapes (National Center for Health Statistics) (18) for people who lived and died in the selected counties for the time periods with fine particulate measurements. After excluding all deaths caused by accidents and other external causes [International Classification of Diseases, 9th Revision (ICD-9) (19), clinical modification codes 800–999], we analyzed the remaining total daily deaths.

**Poisson regression of mortality.** We investigated the association of daily deaths with sources of fine particles separately for each city using Poisson regression in a generalized additive model (GAM) (20,21). That is, in each city we assumed

\[ \log(E(Y_t)) = \beta_0 + \sum S_i(X_{t,i}), \]

where \( Y_t \) is the number of deaths in the city on day \( t \) and \( X_{t,i} \) is the value of covariate \( i \) on day \( t \). GAMs are distinguished by allowing us to use smooth functions \( S_i \) instead of linear terms to control for covariates, such as temperature, that may affect daily deaths in a nonlinear way. Linear functions may be used where appropriate. This approach was introduced for time series of counts in 1994 (22) and is now standard (23,24).

To control for trend and season, we used a locally weighted linear regression (LOESS) smooth function of date with a span of 0.05 (23). For the smooth functions of temperature and dew point temperature, we used LOESS functions with spans of 0.80. Indicator variables for day of the week also were included in the models. This is the identical model used by Schwartz et al. (5) and Laden et al. (8), and more details are provided there. To these models we added a smooth function of the mean PM_{2.5} concentration on the day of death and the previous day, instead of the linear term previously used by Schwartz et al. (5). The smoothing window included 50% of the data, which corresponds to between four and five degrees of freedom for the air pollution relation in each city. Alternatively, we added the estimated mass for each of the source factor scores (in micrograms per cubic meter) simultaneously in the model. That is, the estimate of the mobile source factor is in a model controlling for coal-derived particles, crustal particles, and the other source factors, and vice versa. Because only the particles from traffic showed a strong linear association, and because the exposure ranges for the exposures to coal particles did not overlap sufficiently, we only used a smooth function for the traffic particles and followed Laden et al. (8) in treating the mass from the other sources as linear terms.

**Hierarchical model.** To combine the smooth curves across cities, we applied the approach of Schwartz and Zanobetti (16), as modified by Schwartz et al. (11). In each city, the predicted log relative risk and its pointwise standard error was computed for each 2

<table>
<thead>
<tr>
<th>City</th>
<th>Days</th>
<th>PM_{2.5}</th>
<th>Traffic</th>
<th>Coal</th>
<th>Residual oil</th>
<th>Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston</td>
<td>59</td>
<td>16.5</td>
<td>4.8</td>
<td>8.3</td>
<td>0.5</td>
<td>5/79–1/86</td>
</tr>
<tr>
<td>Knoxville</td>
<td>12</td>
<td>21.1</td>
<td>4.4</td>
<td>6.8</td>
<td>—</td>
<td>1/80–12/87</td>
</tr>
<tr>
<td>St. Louis</td>
<td>55</td>
<td>19.2</td>
<td>2.9</td>
<td>5.6</td>
<td>—</td>
<td>9/79–1/87</td>
</tr>
<tr>
<td>Steubenville</td>
<td>3</td>
<td>30.5</td>
<td>1.5</td>
<td>19.2</td>
<td>0.9</td>
<td>4/79–9/87</td>
</tr>
<tr>
<td>Madison</td>
<td>11</td>
<td>11.3</td>
<td>3.1</td>
<td>4.9</td>
<td>—</td>
<td>3/79–12/97</td>
</tr>
<tr>
<td>Topeka</td>
<td>3</td>
<td>12.2</td>
<td>2.1</td>
<td>7.0</td>
<td>—</td>
<td>9/79–10/88</td>
</tr>
</tbody>
</table>

Figure 1. Overall estimated dose–response relation between total PM_{2.5} and daily deaths in six U.S. cities. The estimate is obtained by combining the estimated smoothed curves in each of the cities, after controlling for weather, season, and day of the week. The shaded area indicates the pointwise 95% confidence intervals at each point. The line shown is a least-squares regression line through the estimated points.
µg/m³ increment in exposure. These estimates are provided by the GAM function in S-plus (MathSoft, Inc., Seattle, WA). To successfully combine data across cities, we need to use a range of exposures that is common to all cities. Because high concentrations of PM$_{2.5}$ were rare, the curves were combined only in the range of 0–35 µg/m³. The first phase of the analysis produced estimated effect sizes (log relative risks) $\hat{Y}_j$ in each city $i$ for each exposure category $j$. A pointwise standard error of the estimate is also estimated by GAM. To produce the combined curve, we regressed these estimates against indicator variables for each city, using inverse variance weighting and allowing for a random variance component to capture heterogeneity in the association across cities. That is, we assumed

$$\tilde{Y}_j \sim N(\beta_0d_1 + \beta_1d_2 + \ldots + \beta_jd_j, V_j + \delta),$$

where $d_j$ are dummy variables for the $j$ exposure levels, $V_j$ is the estimated variance in city $i$ at level $j$, and $\delta$ is the estimated random variance component.

We used the iterative meta-regression approach of Berkey et al. (26) to obtain a maximum likelihood estimate of the random variance component.

The nonparametric smooth functions we use to estimate the shape of the concentration–response relation use four to five degrees of freedom, and it is not clear that the source-specific relations can support so many degrees of freedom, which would entail a total of 20 degrees of freedom for all the PM$_{2.5}$ sources. In our previous report (8), the relation between PM$_{2.5}$ from traffic and daily deaths was estimated with considerably greater precision than for particles from other sources, most of which were not significant. Further, the range of overlap in exposures across cities was lower for coal, crustal, and residual oil factors. Therefore, in our source-specific models, we only modeled the traffic source particles using a nonparametric smooth, while controlling for PM$_{2.5}$ from the other sources using linear terms, as in Laden et al. (8). We then combined the estimated concentration–response relations for traffic particles similarly to what we did for PM$_{2.5}$ from the other sources.

**Results**

Table 1 shows the daily deaths, PM$_{2.5}$ levels, and estimated concentrations of PM$_{2.5}$ from each source. Figure 1 shows the meta-smooth dose–response relation between PM$_{2.5}$ and daily deaths in the six cities. There is no evidence of a threshold, and the relation occurs well below the U.S. Environmental Protection Agency standard of 65 µg/m³ (27). The line shows the least-squares fit of a linear relation through the estimated points.

The next results come from the source component models. These models had a smooth function of PM$_{2.5}$ from traffic and linear functions of PM$_{2.5}$ from the other sources in each city. Figure 2 shows the results when we combined the estimated dose–response curves for traffic particles across the six cities. Again, there is no evidence of a threshold, and the association is essentially linear. If anything, the slope is steeper at lower concentrations. To test the robustness of the association with traffic particles to our method of controlling for particles from other sources, we re-estimated the relationship controlling for smooth functions of the estimated particle mass from other sources, rather than the linear terms. This association is shown in Figure 3 and differs little from that shown in Figure 2. We also fit linear regressions through the points shown on Figures 1 and 2. We obtained a slope of 1.5% increase in deaths per 10 µg/m³ increase in PM$_{2.5}$ and 3% increase in deaths per 10 µg/m³ increase in particles from traffic, which is the same as the results reported by Laden et al. (8). These lines are shown on the figures. This supports the assumption of a linear relationship.

**Discussion**

We have explored the concentration–response relation between PM$_{2.5}$ and daily deaths in six U.S. cities and combined the results to obtain greater stability, while accounting for heterogeneity in response. The population mean curve shows no evidence of a threshold down to the lowest levels of PM$_{2.5}$. In fact, the curve is quite linear over the exposure range from 0 to 35 µg/m³. This is consistent with previous results using a similar methodology but with PM$_{10}$ (10) and black smoke (11) as the exposure metric. In addition, a different methodology, using regression splines, was applied by Daniels et al. (9) to PM$_{10}$ data in different cities. They combined these spline models across 20 cities. Again, the association appeared to be quite linear without any evidence of a threshold. A spline model had previously been applied by Schwartz (22) to the PM$_{2.5}$ data from Boston, with a similar finding. Indeed, the original study of these data by Schwartz, Dockery, and Neas (5) found a significant association when limited to days below 30 µg/m³, with a slightly larger slope. The consistency of the results on two continents, and using different techniques, suggests that this finding is robust. The concentration–response curve seen here for PM$_{2.5}$ is steeper than that previously reported (per µg/m³) for PM$_{10}$ (10). This is consistent with the previous report from this study (5) that coarse mass (the difference between PM$_{10}$ and PM$_{2.5}$) is not associated with daily deaths. We note that Schwartz and Zanobetti (10) demonstrated in simulation studies that measurement error was not likely to distort the shape of the association. Similarly, recent studies of “harvesting” have shown that effect sizes increase rather than decrease when longer lags are taken into account; for example, high
days producing harvesting that mutes the effect on the next high day is unlikely to have distorted the shape of the association.

These results are also biologically plausible. Schwartz (28) pointed out that if thresholds exist in individuals, but there is a distribution of those thresholds among individuals, and if multiple genetic and predisposing illnesses each contributed to the distribution of those thresholds, then by the central limit theorem, the distribution of thresholds should approach a normal distribution. Hence, the population concentration–response curve should approach a cumulative normal curve. But the low-dose end of the cumulative normal curve is linear. To see this, consider that typical death rates in U.S. cities are 8/1,000 per year, or 2 × 10^{-6} per day. The normal range of variation in daily deaths in U.S. cities is a factor of two or less. Hence, the normal range of daily death probabilities in response to all risk factors is from 1 to 3 × 10^{-6}. Figure 4 shows the cumulative normal curve in that range of probabilities, which is quite linear. Because we are clearly in the low-dose regime, in the sense that the exposures to particles are well below the threshold for mortality for most people, this linearity is exactly what would be expected.

Figure 1 also indicates that the association reported here has public health significance. The difference between mean PM_{2.5} concentrations of 10 µg/m³ and 20 µg/m³, which is a difference found between U.S. cities, is associated with about a 1.5% increase in deaths. In a metropolitan area of a million inhabitants, this would amount to about 130 additional early deaths per year, and in the country as a whole, these results indicate that a reduction of 10 µg/m³ would be expected to result in about 36,000 fewer early deaths per year. Although this study does not indicate the extent to which these deaths are brought forward, other studies of the harvesting issue (29–32) suggest that they are considerable.

The association of daily deaths with traffic particles also has no threshold and is somewhat steeper than the association with all PM_{2.5}. This is consistent with the results of Laden et al. (8), except that they used linear terms instead of smooth functions. This study confirms that this association extends to low levels. This result has considerable public policy relevance. Recently, automotive companies have proposed using diesel engines to achieve higher fuel economy in the future. However, diesel engines produce substantially greater emissions of particles and particle precursors such as NOₓ. The present results indicate that such an expansion of diesel engine use in the United States before diesel engines can meet the same particle emission levels as gasoline engines may result in important public health problems. A 1 µg/m³ increase in the concentration of traffic particles in the United States, for example, could be associated with about 7,000 additional early deaths per year in the United States.

**REFERENCES AND NOTES**


