*Environment International*, Vol. 9, pp. 207–220, 1983
Printed in the USA. 

0160-4120/83 $3.00 + .00


# AIR POLLUTION HEALTH EFFECTS IN THE MUNICH METROPOLITAN AREA: PRELIMINARY RESULTS BASED ON A STATISTICAL MODEL

Hans W. Gottinger
Gesellschaft für Strahlen- und Umweltforschung, Institut für Medizinische Informatik und Systemforschung, Ingolstädter Landstrasse 1, 8042 Neuherberg, Federal Republic of Germany

*(Received 2 August 1982; Accepted 7 March 1983)*

Based on a feasibility study, we have recently proposed a multivariate regression model for the impact of air pollution on the mortality of a long-term exposed population. This article proceeds by providing some experimental results, since this model has been fitted to data on a cross sectional basis for 41 census tracts in the Munich metropolitan area. The results obtained so far could be supportive to more detailed epidemiological studies of air pollution which have been conducted in many industrial countries; unfortunately, only relatively few have been conducted in West Germany.

## Introduction and Previous Studies

Based on a feasibility study (submitted to the Ministry of the Environment for the State of Bavaria, Federal Republic of Germany) we have recently proposed a multivariate regression model for the impact of air pollution on the mortality of a long-term exposed population (see H. W. Gottinger, 1981). This model has been fitted to data on a cross sectional basis for 41 census tracts in the Munich metropolitan area (MMA); the present article provides some experimental results. The results obtained so far could be supportive to more detailed epidemiological studies of air pollution which have been conducted in many industrial countries (see Shy, 1979).

Thus far, results from most epidemiological studies have shown an inverse relationship between air quality and mortality. The procedure used in these studies typically involves calculating different mortality rates (or partial correlation coefficients) for populations exposed to different air quality conditions. These studies have been typically controlled for the socioeconomic variables, and possibly for the lifestyle factors, which have an impact on the individual's probability of death. The difficulty with using these types of statistical control techniques is that there are many other factors correlated with air pollution that affect human mortality and that these factors are not adequately controlled by the use of partial correlation coefficients. For example, individuals with relatively high income and education levels tend to reside in less populated areas of relatively better air quality, while the lower income classes tend to be constrained to the more polluted, crowded sections of a metropolitan area. Similarly, occupational mix as a crucial element of "indoor pollution" provides another example. Since income, educational levels, and population density all influence mortality, these relationships cause difficulty in isolating air quality's influence on mortality through the use of simple correlation techniques.

In an attempt to circumvent these limitations, Lave and Seskin, in a series of studies (1972; 1973; 1977), estimated their own relationship using mortality data for 117 Standard Metropolitan Statistical Areas (SMSAs) for the period 1960–1961 on a cross sectional basis, and for 26 SMSAs over a period 1960–1969 on the basis of a time series cross-sectional analysis.

These studies used linear multivariable regression analysis to explain the variance in disaggregated mortality rate according to age, sex, race, and disease category. The independent variables specified were minimum, mean and maximum sulfates; mean, minimum and maximum particulates; percent poor; population per square mile; percent of population 65 yr or older; percent nonwhite; and a standardized population variable of the SMSA population. The results of Lave and Seskin (1973) showed that more than 83% of the variation in the mortality rate was accounted for by the independent variables; however, none of the pollution

variables was statistically significant, although they made a statistically significant contribution as a group. The reason for this could be explained by the problem of multicollinearity among the pollution variables. The results of their linear regressions demonstrate a predominantly direct relationship between levels of pollution and mortality. Their results can be summarized as follows: (a) individuals in the age group 15–44 die primarily from nonpollution related causes; (b) pollution affects persons increasingly with age and/or morbidity; and (c) there exists a cumulative effect of pollution on human mortality.

Another interesting insight is that the influence of particulate matter is greater on (white) males than on (white) females in the younger than 65 yr groups. This result suggests the possibility of relatively higher exposure levels for males as a result of occupational activities.

In an attempt to determine if their results had isolated a true relationship between ambient air pollution and mortality and not a relationship caused by meteorological differentials or personal pollution from home heating sources, Lave and Seskin (1972) added various climatic and home heating variables (e.g., humidity, temperature, and various types of home heating equipment) to their model. The inclusion of these additional variables did not alter the positive nature of the association between air pollution and mortality nor the stability of the air quality coefficients.

Experimentations with alternative functional forms of the regression equations (e.g., log-linear, linear spline, quadratic) have given results not significantly different from the general linear model. Whether the associations between the variables are causal is still an unsettled issue. However, experimental and toxicological results provide supporting arguments for at least the existence of an aggravation effect.

In summary, Lave and Seskin deserve credit of having demonstrated that econometric-statistical modeling studies are likely to be fruitful in the attempt to isolate the aggravation effect of ambient air quality on human mortality.

## Scope and Technical Limitation of Study

The objective of this essay is the estimation of a general statistical model that was presented in Gottinger (1981). A brief description of the model, for completeness' sake, is reproduced in the next section. Using cross-sectional data from the MMA, the various mortality functions are estimated. Specifically, weighted linear regression analysis is used to estimate age-, sex-, and disease-specific mortality functions. Conclusions drawn from this application are then discussed in a forthcoming paper in terms of the physical social benefits (decreases in probability of death) to be expected from improvements in ambient air quality. However, in the present paper we present elasticity estimates indicating the extent of possible impact of the explanatory variables on human mortality, thus establishing a priority ranking of controlling harm. There are various theoretical possibilities to discuss the choice of the regression equations. We will not go into detailed discussion to be found elsewhere in the professional literature (see Lave and Seskin, 1976; Lave and Chappie, 1982), but instead confine ourselves to two points that illuminate the issues.

First, related studies, in particular those by Lave and Seskin (1977), have chosen a linear form for Standard Metropolitan Statistical Areas (SMSAs), and for reasons of comparison it is pertinent to link our results to these reference studies. Lave and Seskin argue that experimentation with alternative specificiations (i.e., functional forms) have given results not significantly different from the general linear model:

> The first of these involved a comparison of the linear model with a log-linear specification. It was found that the elasticities associated with the pollution variables . . . were quite similar to the coefficients of the pollution variables in the linear formulation. When a quadratic model was fit . . . [squared values of the pollution variables as well as interaction term], the results of an $F$-test indicated that the gain in explanatory power was insignificant (Lave and Seskin, 1973).

Second, in looking at functional specifications of dose-response relationships of low-level ionizing radiation, the linear model has been proven to yield a good fit for long-term exposure. Furthermore, a linear model tends to overestimate risks of cumulative long-term exposure, and therefore appears to be particularly suited for regulatory purposes (Land, 1980). The assumption of "conservative" risk assessment by the linear hypothesis appears to be a good initial working hypothesis which subsequently should be checked with sensitivity analysis.

Any attempt to isolate air quality's impact on human mortality must initially recognize that air quality is only one of numerous factors which influence an average individual's probability of death. In multivariable regression analysis of complex dose-response relationships we run easily into problems of spurious correlation, confounding factors and multicollinearity. Recently, various technical tools have been invented for getting rid of multicollinearity, such as ridge regression methods. As discussed by McDonald and Schwing (1973), air pollution variables are often highly correlated with other explanatory variables. They propose ridge regression estimation to circumvent this problem. Ridge regression estimation, however, is only a purely statistical solution to the problem of multicollinearity or "nonorthogonality," and it can lead to arbitrary solutions that are highly sensitive to interpretation of results. While technical solutions might be helpful in certain cases, the situation is a bit more complicated here since we must strike a reasonable balance among

multicollinearity phenomena, data availability of major atmospheric contaminants, and knowledge about those contaminants that have a potential long-run impact upon health. Instead, we pursue some kind of exploratory procedure, closely related to regression diagnostics (Lave and Chappie, 1982), so as to strike a reasonable balance between the various considerations involved.

The choice of candidate variables reflects a theoretical construction of the major relationships, including a selection of socioeconomic factors, climatic variables, and the bulk of air pollution factors. Ideally, all these factors should be controlled in a comprehensive model. In general, it is difficult to get cross-sectional data on "personal habits," such as smoking and dietary habits of the population. Because of lack of knowledge on smoking habits for each census tract in the MMA, we finally decided to drop a smoking index in the final regression model. However, on an exploratory basis, we ran some experiments by including a smoking index for the MMA, derived from the 1978 microcensus on smoking in the State of Bavaria. We found that inclusion of the index would reduce the air pollution coefficients throughout, and in some cases destroy their statistical significance. This kind of explanation would be intuitively acceptable, even on purely theoretical grounds. Since exposure to air pollution is not limited to the atmospheric type of pollution but would also include indoor pollution such as active and passive smoking, and since most people spend 80% of their time indoors, including occupational exposure, it is not surprising that explicit treatment of indoor pollution should have an effect on air pollution coefficients.

However, on the basis of this exploratory procedure thus far, a generalization of this result cannot be suggested for an aggregate model of this type. A similar problem was encountered by Lave and Chappie (1982) who introduced a smoking variable as the "estimated per capita expenditure on tobacco products in each SMSA," which is an average for the entire population of an area and cannot distinguish among nonsmokers, former smokers, light smokers, and heavy smokers. Because of inherent statistical difficulties in representing an explicative smoking variable, only a careful case control study between smokers and nonsmokers, both affected by air pollution, would be able to disentangle the effect of smoking and air pollution.

The major criticisms that have been raised against Lave and Seskin's monumental study (1977) would also apply to this study, and have been satisfactorily answered by Lave and Chappie (1982). The criticisms fall under the following headings:

(1) Aggregate analysis of poor data provides little or no information about the effects of air pollution.

(2) The estimated models omit important variables (e.g., cigarette smoking and diet) or use crude surrogates.

(3) The models failed to treat explicitly the "aggregation problem" arising from the use of data on cities or SMSAs rather than on individuals.

Problem (2) has been treated within this research agenda. Problems (1) and (3) are more intricate since they reveal basically a different philosophy about modelling. The implicit understanding of this philosophy is that assessment of health effects using large aggregate data bases is not feasible or meaningful in terms of policy conclusions. Rather than adopting this attitude, we should better complement such studies by more detailed studies on an individually related basis in order to discriminate between health effects. In particular, it appears that carcinogenic risk assessment models, which have been quite useful for toxicological assessment and regulatory guidelines, could provide additional information on a disaggregated basis; they could therefore be helpful in substantiating claims on changes in population-based mortality and morbidity profiles of aggregate models. It should be borne in mind that the concept of "aggregate model" here is not pushed too far, since the elementary unit for aggregation is the census tract which has been chosen to correspond with the institutional and administrative unit in the MMA. An even higher level of aggregation would be to take the city or the metropolitan area as a basic unit in itself and then to aggregate over city and metropolitan areas (see in this regard the recent review by Ricci and Wyzga, 1980, and Lipfert, 1980).

## Candidate Variables

This model is designed for application to a specific urban area, the MMA, and thus applies to an intraurban analysis of air pollution; in principle, it is also applicable to analyzing interurban mortality differentials.

It is natural to proceed along Lave and Seskin's lines and consider a multivariable regression approach, for the following reasons. Any attempt to provide a reliable estimate of the effects of air pollution on mortality must recognize that other factors which are collinear to air pollution influence the risk of death. Therefore, in order to isolate the effect of air pollution on mortality, it is necessary that these other factors be controlled. The controls are imposed by standardization, i.e., disaggregation of the total mortality rate according to the factors to be controlled. The first of these factors will be demographic influences, controlled through the use of age- and sex-specific mortality rates. In general we can expect mortality differentials with respect to age and sex. Also, it must be recognized that air pollution accentuates the risk of death from certain causes, e.g., bronchitis, emphysema, asthma, etc. These relationships appear well established in previous studies (see Crandall and Lave, 1982). Therefore, deaths should be separated into various causes to enable the isolation of those types which are most likely to be affected by various levels of

air quality. One could also standardize the total mortality rate along social strata; however, this would create a technical ambiguity, since "social strata" then would appear on the side of "dependent variables" as well as on the side of "independent variables" in terms of the surrogate education or income variable.

Toward this end, age- sex-, and disease-specific mortality rates will be used in the present study as dependent variables, and will be aggregated to the total mortality rate.

Incidentally, another advantage of using disaggregated mortality rates is that such variables provide public health indicators of the primary etiological effects of air pollution. In particular, isolating preschool and school-aged children and, to some extent, housewives, avoids technical complications of exposure to pollutants at work.

Referring to the explanatory (independent) variables, the existence of important factors in addition to age, sex, and air pollution that may be collinear to other independent variables and also influence an individual's risk of death should be covered by multivariable regression analysis. These factors include income, education, social class, occupation, place of residence, housing, climate, availability and access to quality medical care, nutritional level, tobacco consumption, sanitation and marital status. This list is by no means exhaustive. Unfortunately, measurement or observation of many of these factors is extremely difficult, and sometimes impossible. Certain proxies or surrogate measures, however, can be used in the absence of direct measurement (Gottinger, 1974).

The following independent variables on the basis of extensive *a priori* knowledge from the relevant literature, have been considered in this prototype model.

*A list of candidate variables*

*Income per head (Y).* In general, individuals with higher incomes tend to consume higher quality nutritional diets which should favorably affect health. A negative association between income and mortality has been demonstrated by researchers detecting class differentials in mortality (see, e.g., Lave and Seskin, 1977). Others, such as Fuchs (1973) and Auster (1969), have shown that after controlling for other factors (e.g., education), income does not indicate a significant negative relationship and sometimes may even exhibit a positive relationship. Income will be used primarily as a proxy for measured socioeconomic variables.

*Education (E).* It is to be expected that "higher levels of education" may be associated with relatively more medical care of preventive stages (Auster, 1969). Moreover, higher educational levels may also be associated with better knowledge of preventive care and healthy life-style habits, and this may contribute to negative associations established between education and mortality.

*Cigarette consumption (C).* The adverse effects of cigarette consumption, through active smoking as well as through passive smoking (contributing to indoor pollution) are well established, and positive associations between cigarette consumption and mortality have been demonstrated in numerous studies (Auster, 1969). Since the carcinogenic risk of smoking may not only shift the disease-specific mortality profile of the population but may also act synergistically with other air pollution contaminants, it may be advisable to model explicitly synergistic effects at a later stage.

*Population density (PD).* Closeness to other individuals can influence one's exposure to various diseases. In particular, we are concerned with the possibility that the populations of the census tracts and consequently their density might be positively correlated with both air pollution and risk of death.

*Air pollution variables.* The air pollution variables are composed of a number of substances that are routinely measured by automated highvolume samplers of monitoring stations in the MMA. Also a limited number of mobile stations routinely collect data on the following substances: sulfur dioxide ($SO_2$), nitrogen oxide ($NO_x$), carbon monoxide (CO), hydrocarbons (CH), dustfall (total particulates), and, in some cases, trace metals (cadmium, beryllium, lead, etc.).

In particular, all evidence from previous air pollution studies suggests that the sulfate-particulate mix, for which there are no direct measurements, is the potentially most dangerous air pollution hazard, more damaging than $SO_2$. In an attempt to construct a realistic model of air pollution variables we are restricted by the availability of on-line data, provided both by monitoring units and by indirect measurements or proxies of those compounds which have been demonstrated to have a significant impact upon health (Landsberg, 1979).

Suspended particulates and/or dustfall should be considered because of the belief that they are of eminent importance for air quality's influence on health due to their almost universal presence and relative magnitude. In order to relate suspended particulates more closely to health effects, it would be advisable to make distinctions in the fineness and restrict oneself to particle fractions in the respirable size range (Katz, 1980). However, not much is known on the cutoff point, and even less is available from the on-line data collections of the monitoring stations. For that reason, we have included all particulate measurements irrespective of their diameter size, and we feel justified by the fact that not much change has occurred in the particulate mix for the years 1973–1977.

Sulfur dioxide should be considered also to test for possible synergistic effects with suspended particulates and/or dustfalls and to act as a proxy for the sulfur-particulate-sulfate complex in addition to the suspended sulfates (SS). Suspended sulfates could also be examined for their possible long-run effects, since the Community

Health and Environmental Surveillance System (CHESS) Studies reveal that short-term exposures to air pollution be blamed for an increased incidence of acute respiratory diseases (see Riggan, 1973). Unfortunately, measurements on suspended sulfates are not sampled by the monitoring stations in the MMA although all evidence suggests that their possible health impact is substantial. Katz (1980) makes clear-cut recommendations on air pollution control strategies:

> . . . since sulfates are mostly secondary pollutants of small particle size and can be transported hundreds of kilometers from an $SO_2$ emission source, measurement techniques must be sufficiently sensitive and accurate to determine their concentration in relation to any future air quality standard and control strategy for sulfates.

Nitrogen oxides (NO) and dioxides ($NO_2$) apparently have an effect on human health. Few studies have been conducted on the toxicity of urban levels of nitrogen dioxide, because the effect of $NO_2$ apparently cannot be separated from those of other pollutants. Another problem is that $NO_2$ could be the major source of indoor pollution, since it is commonly generated by tobacco smoking and gas cooking and heating. Nitrogen oxide in connection with hydrocarbons (CH) is the major source of photochemical smog. Because of complex chemical reactions in the atmosphere and collinearity with other air pollution contaminants, we omitted a separate analysis of $NO_x$.

*Meteorological variables.* Turning now to the meteorological complex of variables, we restrict ourselves to include only two: the number of days (per year) with precipitation greater than 5 mm (NM) and number of days with maximal temperature less than or equal to 0 °C (TEMP). Other meteorological variables could be potential contributors to air pollution and mortality patterns, but we limit ourselves to those which are the most readily available and which appear to be of universal validity. Although climate has been recognized as an important factor in mortality, studies on a potential cause–effect mechanism have not been developed to the point where it can be determined which variables are of major importance.

## The Reference Model

By considering the initial variable selection and by hypothesizing that for long-term, chronic exposure of pollutants the relationship is assumed to be linear, the following general mortality function is proposed:

$$M_{asci} = b_0 + b_1 Y_i + b_2 E_i + b_3 C_{asi} + b_4(\mathrm{PD})_i + b_5(\mathrm{SO_2})_i + b_6(\mathrm{CO})_i + b_7(\mathrm{TP})_i + b_8\mathrm{TEMP} + b_9\mathrm{NM} + e_i.$$

The subscript $i$ defines census tracts in the MMA. For census tracts, we used the most elementary political units, the districts which in the MMA are of varying size, and which are covered by at most one stationary or mobile monitoring unit. If this requirement is not fulfilled in one case or another, the arithmetical average of at least two adjacent stationary monitoring units is used as an estimate. This procedure appears questionable from an aerometric perspective. However, since the census tracts of the MMA appear relatively small in diameter and since quite a few of them had stationary or mobile monitoring units, the computations could be taken as first approximation. Implicitly, it comes close to assuming that the air pollutants will distribute uniformly according to a Gaussien plume of diffusion along the line between two adjacent areas. $M_{asci}$ denotes the mortality rate for disease $c$, age group $a$, and sex $s$ in tract $i$. The age groups are specified for three classes: years $<45$, 45–65, $>65$. The disease complex is organized according to the International Classification of Diseases (ICD) code where those chronic diseases are selected which appear to have a proven linkage to air pollution variables.

The reason for the decomposition of the global (all-cause related) mortality rate into an air pollution related group $c$ and the complement of the group $\bar{c}$ is related to the fact that long-term exposure of the population by ambient contaminants has a significant impact on the risk profile of specific diseases. The global mortality rate is then $M_{asi} = M_{asci} + M_{as\bar{c}i}$. The overall or total mortality rate on a communal basis in the MMA is the aggregate over the age, sex and census tracts, denoted by $M$. The split into age groups $<45$, 45–65, $>65$ permits the investigation of the aggravation effects of smoking and occupational mix in addition to air pollution. It can be assumed that in the first age group these aggravation effects play a minor role because of their delay on mortality and morbidity.

Throughout the study, generalized least-squares (GLS) method was applied. Mortality rates have been taken from the Statistical Office of the MMA, which are disaggregated according to the 41 census tracts (Bezirke), standardized by age and sex, but unfortunately not by disease categories. The mortality rates were recomputed according to almost standard demographic procedures, for a period of 5 yr (1974–1978), to get an average mortality rate (see Appendix A). Disease-specific mortality rates for each census tract were obtained by a "tricky" procedure in which the age-sex mortality rates were corrected mainly for suicides, accidents, and other non-pollution related categories on the basis of age and sex in each census tract, since there are nonchronic mortality causes such as infectious diseases and sudden deaths in early infancy (see Harriss, 1979).

Some problems arising from migration trends to and from the MMA are treated in such a way that for the observation period concerned (1973–1977) it is assumed that people moving into and away from the MMA on average have been exposed to a comparable level of air pollution. This appears to be a heroic assumption, in

fact, even if not true; it has, however, an insignificant impact on the statistical results.

## Data Sources and Data Availability

The procedures used to estimate the variables considered in the prototype model, the units of measurement, and the data sources are presented below. Census tracts were used as the primary unit of measurement because such divisions are believed to provide a more accurate assessment of the individual socioeconomic factors. The age-, sex-, and disease-specific mortality rates, e.g., $M_{asci}$, were based on deaths occurring in the MMA during the 5-yr period 1974–1978, allowing for a 1-yr response period after exposure of the population to air pollution in the period 1973–1977. The justification for this is mainly theoretical plausibility, to preserve a quasicausal relationship but in a delayed form. Indeed, the delayed response period of 1 yr may be too short for certain diseases (e.g., cancer or obstructive lung disease), and too long for others (cardiovascular diseases). However, in the case of chronic and cumulative exposure it is essential that mortality profiles and explanatory variables are relatively stable on the average. It appears that longer delays do not significantly affect the results, basically because there is no discernible structural change in the mortality profile. Further investigations are necessary.

The rationale for using a 5-yr average mortality rate is that it will help reduce the variance caused by small and differing sample sizes (population of the census tracts). This is an important consideration since the stability of these mortality rates, rather than their absolute size, is essential. For illustrative purposes, consider a census tract with 20 individuals, one of whom dies during our 5-yr study. In this case the 5-yr average mortality rate will be 1/100, although the yearly rates will be highly unstable, ranging from 1/20 to 0.

The 1973–1977 per capita income in census tract $i$ is calculated from publications of the Munich Statistical Office regarding the Munich Gross Domestic Product (for 1976) and the distribution of household incomes (for 1980). The percentage of the population, 25 yr old and older, in census tract $i$ with at least high school (Gymnasium) education was calculated from data available from the same office on the educational output of high schools and equivalent school types.*

For experimental purposes, the estimated tobacco consumption for age group $a$, sex $s$, for a few census tracts was derived from the microcensus in which smoking habits of the Bavarian population were investigated on the basis of a 1% percent random sample. From the microcensus data, crude information can be derived about the association of smoking habits with demographic and social factors; thus, at least in principle, a proxy can be established for $C_{asi}$. Since age, sex, and social class constitute the basic information being linked to smoking, an attempt is made to extrapolate smoking habits based on this information for each population in the census tract. The estimation procedure establishing this association assumes that the MMA is typical for the State of Bavaria as a whole.

Three alternative measures of population density (crowding) were used. The first of these variables was population per hectare in census tract $i$ $(P/\text{ha})_i$. This figure is calculated by dividing the total 1975 population of census tract $i$ by the area of tract $i$ in hectare. The second proxy for a crowding variable, population per room $(P/R)_i$ in census tract $i$, is calculated by dividing the population in $i$ by the total number of rooms occupied by the individuals in tract $i$. The third, most pertinent, crowding variable proxy, population per residential area $(P/\text{RA})_i$ in tract $i$, is calculated by dividing the 1975 census population of tract $i$ by the total residential area (living space) in tract $i$.

Referring now to the air pollution variables, we select those which are continuously monitored by the 16 fixed monitoring stations in the MMA, supplemented by occasional measurements of 20 mobile monitoring units, under the supervision and legal control of the Bavarian Office for Environmental Protection (Bayerisches Landesamt für Umweltschutz).

The on-line measurements include $SO_2$, total particulates, $NO_x$, and CO; sample measurements only are available for CH and trace metals. The 5-yr arithmetic mean of annual averages of total particulates in $\mu g/m^3$, ppm, or ppt, for the calendar years 1973–1977, is calculated for each census tract using the following procedure:

(a) Calculate the 5-yr average of annual arithmetic means by monitoring station. Use the average values for a specific individual station to estimate missing values for that station.

(b) Locate the individual monitoring stations on a map of the MMA and identify their individual census tracts and its boundaries. Estimate values for the remaining points in the MMA by using an interpolation procedure. This procedure involves calculating the weighted average of slopes from values of nearby data points and is modified to consider distance and direction.

(c) Plot these interpolated values on a map of the MMA in order to facilitate the estimation of average values for each census tract.

The original data on total particulates are collected at the individual air monitoring stations operated in the MMA.

*A complete list of data sources, statistical material and other documentary evidence of these and the subsequent descriptions (in German) are available upon request from the author. Therefore they have not been included in the references.

The 5-yr (1973–1977) arithmetic mean annual averages of sulfur dioxide in ppt/24 h and carbon monoxide (CO) for the calendar years 1973–1977 are calculated for each census tract using the same procedure applied to total particulates. Among the meteorological variables, TEMP and NM were recalculated from data made available by the Munich Meteorological Survey.

## Estimation Problems

The general model is estimated using the variables described previously. More specifically, 12 age,- sex-, and disease-specific linear mortality functions are estimated using weighted regression analysis. The rationale for using weighted regression analysis is to correct the problems of nonconstant variance of the error term (heteroscedasticity). The existence of heteroscedasticity violates the assumptions of ordinary least-squares regression analysis, thereby introducing biases into the estimated results.

We follow a procedure suggested by Smith (1968) for elimination of this problem in cross-sectional analysis. Instead of minimization of the quadratic error $\Sigma e_i^2$ the regression equation system, where $e_i$ indicates the difference between observed and estimated mortality rate, it is suggested to minimize the expression.

$$\Sigma e_i^2/(s^2/N_i),$$

or, if $s$ is constant, to minimize

$$\Sigma e_i^2 N_i = \Sigma(e_i\sqrt{N_1})^2.$$

In other words, this means the error should be weighted by a factor equal to the square root of the sample size. Following this correction the linear weighted regression analysis resulted in the best estimates based on the following three criteria:

(a) The highest (corrected) coefficient of determination $\bar{R}^2$ found to be significant by the corresponding $F$ statistics,
(b) regression coefficients that are statisticaly significant ($t$ test),
(c) a residual pattern which best supports the assumptions that the error terms are independent, and have zero mean and constant variance.

For example, all the independent variables associated with male mortality in the over-65 age group in each observation are multiplied by the square root of the male population over 65 in the corresponding census tract.

In the regressions which follow, several candidate variables are excluded from the equations. Among the three alternative crowding variables we opted for $P/(\text{ha})_i$, for reasons of convenience, which in the high density MMA apparently is not much at variance with $P/(\text{RA})_i$. The selection of climatic variables is based on statistical significance requirements; TEMP and NM turn out to be most important. In the case of $Y$, the decision not to include this variable is based on the desire to avoid the effects of multicollinearity. Specifically, this variable was highly correlated with $E$. Furthermore, it can be shown that the estimates for $C$ are extremely unreliable. In some cases, however, it appears evident that explicit consideration of the $C$ variable could destroy statistical significance of the air pollution coefficients. Therefore, for such cases, the results should be interpreted more cautiously. Despite initial considerations practical experience led us to omit the $C$ variable. To facilitate the discussion, the means and standard deviations of the variables included in the various functions are presented in Table 1.

Table 1. Means and standard deviations of variables included in equations.

| Variable | Mean | Standard Deviation |
|---|---|---|
| Percent of adult population (>25 years of age) with high school (gymnasium) education or equivalent ($E$) | 22.06 | 8.12 |
| Total particulates (TP) ($\mu g/m^3$) | 230.73 | 18.65 |
| Sulfur dioxide ($SO_2$) (ppt/24 h) | 38.36 | 9.83 |
| Carbon monoxide (CO) (ppm/24 h) | 5.80 | 1.84 |
| Number of days with precipitation >5 mm (NM) | 85.16 | 3.72 |
| Number of days with max. temperature ≤0 °C (TEMP) | 24.75 | 2.34 |
| Population per hectare ($P$/ha) density | 42.00 | 32.36 |

## General Results

The results of the weighted linear regression analysis using the final set of independent variables regressed on the various age-, sex-, and disease-specific mortality rates are presented in Table 2. The figures in parentheses represent the student $t$-values associated with each coefficient. The statistical significance of each variable's coefficient is indicated in appropriate fashion. The inclusion of additional "candidate" independent variables for each mortality function did not significantly increase the equation's overall explanatory power ($\bar{R}^2$). The percent of the variation in the dependent variables explained by the individual regressions presented in Table 2 range from 21% for female pollution-related causes of death in the <45 age group to over 74% for male overall mortality in the 65-and-over age group.

Table 2 consists of three main parts. Table 2A presents the results of applying the general linear model to overall (global) mortality. The signs of the coeffi-

cients for all variables are as expected with the exception of $SO_2$ for males younger than 45. The absolute value of all coefficients increases with age as does the $\bar{R}^2$. The increase in $\bar{R}^2$ is probably the result of both the inverse relationship between accidental deaths and age and the direct relationship between accidental deaths and age and the direct relationship between sample size (number of deaths) and age. The unexpected negative signs on air pollution variables could occur because younger people die primarily from nonpollution-related causes.

In order to test this hypothesis, separate mortality functions are estimated for pollution-related (Table 2B) and nonpollution-related (Table 2C) causes of death using the same model as for overall mortality. This procedure allows the testing of this hypothesis since the coefficients of the variables in the overall mortality functions are simply the sum of the coefficients from the pollution-related and nonpollution-related mortality functions.

For the pollution-related causes of death, all air pollution variables have the expected sign, thus indicating that air quality reinforces these types of illnesses regardless of age. This would still hold true if other specifications of the model would have been adopted. Remember, we chose a conservative procedure. It is also noteworthy that the coefficients increase with age, indicating that possibly the absolute effect of air pollution is greater on the elderly or that there exists a cumulative

Table 2. Mortality function estimates (in deaths per 100,000) (1974–1978), Munich metropolitan area (MMA).[a]

| Variable | <45 Yr | | 45–65 Yr | | >65 Yr | |
|---|---|---|---|---|---|---|
| | Male | Female | Male | Female | Male | Female |
| A. Global Mortality Functions | | | | | | |
| Constant | 288.889 | 218.715 | 1613.719 | 693.977 | 2722.679 | 2841.319 |
| Percent of adult population (>25 years of age) with high school education or equivalent (*E*) | −1.497 (5.361)[b] | −0.624 (2.953)[b] | −15.382 (7.350)[b] | −5.962 (5.822)[b] | −40.725 (6.413)[b] | −11.785 (2.254)[b] |
| TP ($\mu g/m^3$) | 0.306 (1.401) | −0.679 (0.447) | 4.152 (2.404)[b] | 2.665 (3.252)[b] | 12.009 (2.046)[b] | 12.039 (2.445)[b] |
| $SO_2$ (ppt/24 h) | −0.481 (1.097) | 0.061 (0.201) | 1.769 (0.512) | 0.402 (0.253) | 1.484 (0.136) | 2.321 (0.266) |
| CO (ppm/24 h) | 0.386 (1.269) | 0.254 (0.986) | 1.468 (0.612) | 1.076 (0.514) | 1.695 (0.693) | 1.895 (0.795) |
| Number of days precipitation >5 mm(NM) | 2.713 (4.825)[b] | 1.718 (4.232)[b] | 16.595 (3.509)[b] | 5.446 (2.342)[b] | 104.790 (6.345)[b] | 52.899 (3.900)[b] |
| Number of days max. temp. ≤0 °C (TEMP) | −1.715 (1.584)[b] | −1.415 (1.846)[c] | −3.247 (0.349) | 2.364 (0.525) | −58.365 (1.822)[c] | −36.115 (1.388) |
| Population per hectar (*P*/ha) | 0.027 (0.899) | 0.010 (0.456) | 0.689 (2.902)[b] | 0.233 (2.243)[b] | 0.372 (0.620) | 0.307 (0.661) |
| $\bar{R}^2$ | 0.486 | 0.381 | 0.612 | 0.620 | 0.742 | 0.698 |
| B. Pollution-related disease-specific mortality functions | | | | | | |
| Constant | 22.002 | 68.982 | 1128.087 | 354.199 | 1752.255 | 2077.786 |
| Percent of adult population with high school education (*E*) | −0.431 (3.945)[b] | −0.067 (0.745) | −9.984 (6.619)[b] | −4.492 (6.891)[b] | −28.134 (5.651)[b] | −9.165 (2.294)[b] |
| TP ($\mu g/m^3$) | 0.183 (2.145)[b] | 0.057 (0.880) | 2.698 (2.167)[b] | 1.364 (2.616)[b] | 0.746 (2.117)[b] | 8.431 (2.241)[b] |
| $SO_2$ (ppt/24 h) | 0.026 (0.152) | 0.76 (0.592) | 1.538 (0.618) | 0.321 (0.316)[b] | 2.457 (0.286) | 5.111 (0.766) |
| CO (ppm/24 h) | 0.244 (0.212) | 0.182 (0.151) | 1.192 (0.924) | 0.746 (0.478) | 1.423 (1.075)[b] | 1.560 (1.325) |
| Number of days precipitation >5 mm (NM) | 0.155 (0.704) | 0.116 (0.672) | 11.542 (3.387)[b] | 3.184 (2.150)[b] | 71.915 (5.555)[b] | 37.346 (3.604)[b] |
| Number of days max. temp. ≤0 °C (TEMP) | 0.500 (1.181) | −0.335 (1.032) | −4.283 (0.639) | 1.828 (0.638) | −38.055 (1.515) | −25.890 (1.302) |
| Population per hectar (*P*/ha) | 0.006 (0.536) | 0.015 (1.663)[c] | 0.361 (2.123)[b] | 0.152 (2.303)[b] | 0.080 (0.170) | −0.265 (0.746) |
| $\bar{R}^2$ | 0.366 | 0.210 | 0.554 | 0.566 | 0.703 | 0.669 |

Table 2. (Continued)[a]

| Variable | <45 Yr Male | <45 Yr Female | 45-65 Yr Male | 45-65 Yr Female | >65 Yr Male | >65 Yr Female |
|---|---|---|---|---|---|---|
| C. Nonpollution-related cause-specific mortality functions | | | | | | |
| Constant | 266.867 | 149.732 | 485.631 | 339.779 | 970.423 | 763.535 |
| Percent of adult population with high school education ($E$) | −1.066 (4.413)[b] | −0.557 (2.870) | 5.399 (5.904)[b] | −1.470 (2.537)[b] | −12.591 (5.560)[b] | −2.620 (1.391) |
| TP ($\mu g/m^3$) | 0.123 (0.650) | 0.015 (0.893) | 1.454 (1.926)[c] | 1.301 (2.804)[b] | 2.264 (1.081)[b] | 3.608 (2.035)[b] |
| $SO_2$ (ppt/24 h) | −0.507 (1.337) | −0.015 (0.055) | 0.231 (0.152) | 0.081 (0.090) | −0.972 (0.249) | 2.791 (0.888) |
| CO (ppm/24 h) | 0.124 (0.128) | 0.072 (0.065) | 0.276 (0.185) | 0.330 (0.289) | 0.272 (0.202) | 0.335 (0.265) |
| Number of days precipitation >5 mm (NM) | 2.558 (5.258)[b] | 1.602 (4.296)[b] | 5.052 (2.445)[b] | 2.262 (1.718)[b] | 32.875 (5.583)[b] | 15.553 (3.184)[b] |
| Number of days max. temp. ≤0 °C (TEMP) | −2.215 (2.364)[b] | −1.079 (1.533) | 1.037 (0.255) | 0.536 (0.210) | −20.309 (1.778)[c] | −10.225 (1.091) |
| Population per hectar ($P$/ha) | 0.021 (0.797) | −0.005 (0.272) | 0.324 (3.140)[b] | 0.081 (1.373) | 0.292 (1.363) | −0.042 (0.253) |
| $\bar{R}^2$ | 0.432 | 0.387 | 0.520 | 0.546 | 0.614 | 0.570 |

[a] Figures in the parentheses represent student $t$-values associated with each coefficient.
[b] Significant at the 0.05 probability level (two-tailed test).
[c] Significant at the 0.10 probability level (two-tailed test).

effect of exposure to air pollution. To achieve a clearer interpretation of the results obtained, each independent variable must be examined.

The remainder of this section analyzes the estimated coefficients for the various independent variables in order to derive a better understanding of the theoretical and policy implications associated with responses in human mortality resulting from differences in air quality.

The percentage of the adult population with gymnasium education is inversely related to all mortality rates. Using the pollution-induced mortality function for males 65 yr and over as an example, an increase of 1% of the adult population with a gymnasium education will lead to a decrease of approximately 28 pollution-induced deaths per annum for every 100,000 males in this age group.

The coefficients of $E$ are highly significant (as indicated by the $t$ statistics in parentheses below the coefficient) in all mortality functions except for the younger-than-45 female pollution-induced causes and the 65-and-over female nonpollution-induced causes. In these cases, the coefficients of $E$ are significant at the 0.50 and 0.20 probability levels (two-tailed tests), respectively. The relative insignificance of this variable in the younger-than-45 case can be attributed to the comparatively small sample on which the mortality rate is based. Specifically, as illustrated in Table 3, only 400 females in the younger-than-45 group (20% of all deaths in this age-sex group) died from pollution-related causes during the 1974–1978 period. It may be concluded that, in general, the effect of education is significantly different from zero and inversely related to mortality, thereby suggesting that more education results in increased quality and/or quantity of medical care at the preventive stages of an illness. It is also interesting to note that the absolute size of $E$'s coefficient is larger (indicating a lower probability of death, *ceteris paribus*) for pollution-related causes than for nonpollution-related causes regardless of sex after age 45. This result may be attributable to favorable occupational and/or residential exposure to air pollution on the part of more educated individuals.

The coefficients of $SO_2$ are not significant for any of the mortality functions. This result does not imply that $SO_2$ has no impact on mortality. In fact, this result may indicate that $SO_2$ requires a carrier (e.g., particulate matter) in order to enter the body and exhibit its deleterious effects, or that concentrations found in the MMA during the 1973–1977 period were not intense enough to induce a statistically significant effect. It should be noted that when a multiplicative interaction term between $SO_2$ and total particulates was used in the regression analysis, the coefficient of this synergistic variable was predominantly positive and significant at the 0.15 probability level (two-tailed test).

Even though the estimated $SO_2$ coefficients are statistically insignificant, they warrant further discussion since $SO_2$ has been widely considered a major contributor to the adverse health effects of air pollution. $SO_2$ exhibits the expected positive relationship to overall and pollution-induced mortality, although $SO_2$ is often inversely related to nonpollution-induced mortality

Table 3. Mortality distribution in the MMA according to age, sex, and cause. Deaths during 1974–1978. (Figures in parentheses are percentage of column total.)[a]

| Deaths | <45 Yr | | 45–65 Yr | | >65 Yr | |
|---|---|---|---|---|---|---|
| | Male | Female | Male | Female | Male | Female |
| Overall | 3,305 | 2,003 | 11,147 | 6,346 | 21,604 | 22,096 |
| | (100) | (100) | (100) | (100) | (100) | (100) |
| Pollution-related | 707 | 400 | 7,476 | 3,212 | 15,461 | 15,963 |
| | (21) | (20) | (67) | (50) | (71) | (72) |
| Nonpollution-related | 2,598 | 1,603 | 3,671 | 3,224 | 6,143 | 6,133 |
| | (79) | (80) | (34) | (50) | (29) | (28) |

[a] The smaller number of entries (deaths) in one cell is 8 (in the case of female pollution-related deaths <45 yr old), the largest number of entries (deaths) in one cell is 690 (in the case of male pollution-related deaths >65 yr old).

(four of the six coefficients have negative signs). No *a priori* rationalization exists for a positive relationship between $SO_2$ and nonpollution-induced causes of death. Consequently, the predominance of negative signs is not unexpected.

On the other hand, using the pollution-induced mortality function for males 65 and over as an example, an increase of average annual $SO_2$ concentrations by one ppt/24 h will result in an increase of approximately 2.5 pollution-induced deaths per year for every 100,000 males in this age group. It should also be recognized that the size of the $SO_2$ coefficients increase with age for all pollution-induced mortality functions and that these $SO_2$ coefficients are constantly larger for pollution-induced mortality functions than for the nonpollution-induced mortality functions. As Table 3 indicates, the increase in the size of the coefficients with age is to be expected, since the number of individuals dying from pollution-induced causes increases with age. The probable explanations for these results are either that the ability of the body to absorb $SO_2$ influence decreases with age, or there exists a cumulative influence of $SO_2$ on mortality.

Total particulate matter predominantly exhibits the expected positive relationship to all mortality rates although total particulates (TP) is inversely related to overall and nonpollution-induced mortality for females in the younger-than-45 age group. The coefficients are generally statistically significant for most of the mortality functions. Using the pollution-induced mortality function for elderly males as an example, an increase of average annual TP concentrations by one $\mu g/m^3$ will result in an increase of approximately 10 pollution-induced deaths per year for every 100,000 males in the 65-and-over age group. The size of the TP coefficient is larger for the pollution-induced mortality functions than for the nonpollution-induced mortality functions. The coefficients of TP increase with age for all mortality functions, not just pollution-induced as is the case with $SO_2$'s coefficients. The observed sex-differential in the size of the TP coefficient suggests the possibility of relatively higher exposure levels at work for males, compared to exposure levels experienced by females in the residential environment.

In general, it can be concluded that the effect of TP is significantly different from zero and positively related to mortality, especially pollution-induced mortality. Similar results are obtained for carbon monoxide; they suggest a cumulative impact of carbon monoxide leading to adverse health effects with increasing age, independent of sex, but strongly in pollution-induced mortality.

Turning now to the climatic set of variables, we see that the number of days with substantial precipitation is directly related to all mortality rates. The coefficients are highly significant in all mortality functions except for the younger-than-forty-five age pollution-induced causes. In these cases, the coefficients are significant at the 0.50 probability level. It should also be noted that the absolute size of the coefficients increases with age and is consistently larger for males in each age-cause mortality function. These results indicate that the elderly and males are more severely impacted by rainfall.

The number of days with temperature less than 0 °C is negatively related to mortality in most of the mortality functions estimated and positively related to the remainder of the mortality functions. It appears that these results, together with the fact that TEMP is only significantly different from zero in 50% of the cases examined, can only be explained by asserting that cold weather influences various age-sex groups differently.

Population per hectare in each census tract generally exhibits a positive relationship to mortality, although *P*/ha is inversely related to all female over-65 mortality rates as well as nonpollution-induced female mortality in the younger-than-45 age group. Since the *a priori* expectation was for a positive relationship between population density and mortality, these negative coefficients are significantly different from zero at the 0.40 probability level (two-tailed test). All *P*/ha coefficients in the 45–64 age group are positive and significantly different from zero at the 0.20 probability level. Thus, it can be tentatively concluded that population density generally

increases the probability of death for the nonretired age groups (younger than 65), possibly caused by increasing exposure to a variety of diseases.

## Elasticity Estimates

An interesting way of viewing the regression results in all parts of Table 2 is to represent them in terms of the elasticities of the independent (explanatory) variables. Specifically, elasticity estimates facilitate the comparison of effects from a single explanatory variable—*ceteris paribus*—on a variety of dependent variables (mortality rates). In other words, the elasticity of the explanatory variable on the dependent variable is the change of the logarithm of the explanatory variable divided by the change of the logarithm of the dependent variable. The three parts of Table 4 correspond to the parts of Table 2, and present the elasticities (calculated at the mean) of each explanatory variable in the various age-, sex-, and disease-specific mortality functions. As an illustration, referring to the earlier results in Table 2B, they may be reinterpreted as estimating that a 100% increase in each of the explanatory variables will produce the following percentage changes in annual elderly male pollution-related mortality rates:

Percent of adult population (>25 years of age) with high school education ($E$) = −6.5;
Total particulates (TP) = +4.8;
Sulfur dioxide ($SO_2$) = +0.5;
Carbon monoxide (CO) = +0.25;
Number of days with precipitation >5 mm (NM) = +27.2;
Number of days with maximum temperature ≤0 °C (TEMP) = −6.2;
Population per hectare ($P$/ha) = +0.1.

It can also be seen from Table 4B and 4C that the effects of TP on the various pollution-related mortality rates are generally greater than the effect of TP nonpollution-related mortality rates. In fact, TP have their greatest relative effect on males, while the effect of TP on elderly males is approximately the same as on elderly females: from the estimates in Table 4 it can also be concluded that TP have a larger relative effect on all pollution-related mortality rates than $SO_2$ even if the estimated $SO_2$ coefficients were statistically significant.

A possible explanation for this result is that some of the particulate matter located in the MMA may be independently and inherently harmful (e.g., asbestos) or that the particulate matter acts as a carrier of other hazardous elements.

Finally, the weight of the evidence from the MMA suggests that long-term low-level dosages of air pollution, measured as TP and $SO_2$, do in fact have exacer-

Table 4. Elasticity estimates.

| Variable | <45 Yr Male | <45 Yr Female | 45–64 Yr Male | 45–64 Yr Female | >65 Yr Male | >65 Yr Female |
|---|---|---|---|---|---|---|
| A. Explanatory variables with respect to global mortality functions | | | | | | |
| $E$ | −0.026 | −0.015 | −0.047 | −0.041 | −0.064 | −0.019 |
| TP | 0.009 | −0.035 | 0.027 | 0.040 | 0.040 | 0.042 |
| $SO_2$ | −0.005 | 0.001 | 0.003 | 0.002 | 0.001 | 0.002 |
| CO | 0.003 | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 |
| NM | 0.074 | 0.066 | 0.080 | 0.060 | 0.263 | 0.137 |
| TEMP | −0.019 | −0.023 | −0.006 | 0.01 | −0.060 | −0.039 |
| $P$/ha | 0.002 | 0.001 | 0.008 | 0.007 | 0.003 | −0.002 |
| B. Explanatory variables with respect to pollution-related disease-specific mortality functions | | | | | | |
| $E$ | −0.090 | −0.005 | −0.044 | −0.061 | −0.067 | −0.020 |
| TP | 0.082 | 0.009 | 0.025 | 0.004 | 0.046 | 0.040 |
| $SO_2$ | 0.003 | 0.004 | 0.004 | 0.003 | 0.004 | 0.007 |
| CO | 0.095 | 0.026 | 0.003 | 0.004 | 0.002 | 0.001 |
| NM | 0.051 | 0.014 | 0.080 | 0.069 | 0.274 | 0.132 |
| TEMP | 0.069 | −0.017 | −0.012 | 0.015 | −0.060 | −0.038 |
| $P$/ha | 0.005 | 0.005 | 0.006 | 0.008 | 0.001 | 0.002 |
| C. Explanatory variables with respect to nonpollution-related disease-specific mortality functions | | | | | | |
| $E$ | −0.020 | −0.018 | −0.054 | −0.021 | −0.058 | −0.016 |
| TP | 0.005 | 0.090 | 0.031 | 0.042 | 0.022 | 0.017 |
| $SO_2$ | −0.006 | −0.006 | 0.001 | 0.001 | −0.003 | 0.010 |
| CO | 0.005 | 0.004 | 0.001 | 0.003 | 0.005 | 0.004 |
| NM | 0.078 | 0.087 | 0.080 | 0.015 | 0.241 | −0.148 |
| TEMP | −0.024 | −0.024 | 0.006 | 0.005 | −0.062 | −0.041 |
| $P$/ha | 0.002 | −0.001 | 0.013 | 0.005 | 0.005 | −0.001 |

bating effects on certain (pollution-related) causes of death. These effects, while small in absolute terms, are statistically significant for TP although statistically insignificant for $SO_2$ when $SO_2$ is considered as a separate independent variable.

Therefore, based on the results obtained for the MMA, it may be expected that an improvement in ambient air quality will produce social benefits in the form of decreased probabilities of death. These potential mortality reductions, in addition to the numerous other potential costs and benefits, must be further evaluated in order to determine the socially optimum level of air quality.

From the point of view of a cost-effectiveness strategy of implementing pollution controls by the regulatory agency (e.g., the Environmental Protection Office) we present an illustrative model in Appendix B which links the ranking induced by the elasticity estimates to the comprehensive goal of cost-effectiveness of pollution control.

## Highlights of Policy-Related Findings

1. The results are overall in agreement with the extensive studies conducted by Lave and Seskin (1977) for interurban mortality patterns in the United States, although our coefficients are in general lower than Lave and Seskin's results.

2. There appears to be no doubt that cumulative exposure of air pollution contaminants such as $SO_2$, TP, and CO provides a significant health risk in particular to the elderly ill and less educated section of the exposed population. The combined action of all air pollution variables constitutes a health risk that is statistically significant for all age groups regardless of sex. If indoor pollution such as passive and active smoking is separately treated, the air pollution coefficients are reduced and in some cases become statistically nonsignificant, yet even then the results suggest that air pollution by itself constitutes a non-negligible aggravating factor.

3. It appears by sketchy preliminary results that lifestyle factors (exercises, dietary habits) that individually are more likely to be controlled constitute a stronger contributing factor to general health maintenance than air pollution factors at current average MMA pollution levels (except perhaps for the ill and elderly).

4. For pollution-related mortality functions, the results indicate that air quality reinforces these types of illnesses regardless of age; furthermore, the absolute effect of air pollution is greater on the elderly (or there exists a cumulative effect of exposure to air pollution).

5. The percentage of the adult population with high school (gymnasium) education or equivalent is inversely related to all mortality rates, thereby suggesting that more education results in increased quantity and/or quality of medical care at the preventive stages of an illness. The same observation on the educational variable has been made by Lipfert (1980) and the result should be interpreted with caution. "The result indicates an improvement in life expectancy for older people in locations that have well-educated people, not necessarily that college graduates live longer."

6. Although the coefficients of $SO_2$ are not significant for any of the mortality functions, it should be noted that when a multiplicative interaction term between $SO_2$ and total particulates was used, the synergistic effect was predominantly positive and significant at the 0.15 probability level (two-tailed test). The result indicates that $SO_2$ requires a carrier (e.g., particulate matter) to form a sulfur-particulate mix causing more serious health effects.

7. The observed sex-differential in the size of the TP coefficient suggests the possibility of relatively higher exposure levels at work for males compared to exposure levels experienced by females in the residential environment.

8. The absolute size of the air quality coefficients in the pollution-related mortality functions are generally twice as large for males as for females in the younger-than-65 age group. For the 65-and-over age group, however, the coefficients are approximately equal for all males and females.

9. The absolute size of the precipitation coefficients increases with age and they are consistently larger for males in each age-cause mortality function, indicating that the elderly and males are more severely affected by rainfall.

10. The estimated air pollution coefficients indicate that at the mean, a 10% reduction in total particulates will lead to between a 0.04% and a 0.82% reduction in pollution-related mortality. It should be recognized that even if the $SO_2$ coefficients were statistically significant, a similar 10% reduction in $SO_2$ would only result in between a 0.03% and a 0.07% reduction in pollution-related mortality. In essence, this result implies a relatively greater importance of TP compared to $SO_2$.

11. It can be tentatively concluded that population density (residential crowding) generally increases the probability of death for the nonretired age (younger than 65), possibly caused by increasing exposure to a variety of diseases.

## Topics for Further Research

A research agenda for future work comprises various new directions to fill some remaining gaps of the present analysis:

1. Explicit modelling of occupational exposure and consideration of life-style factors in relation to atmospheric air pollution variables.

2. Reconsideration of various types of total particulates, e.g., those of diameter size less than 0.7 $\mu$m,

because those of larger size are likely to be excluded from penetration into bronchi. (Particle size may vary over the years and the health impact of TP virtually depends on its structure and size.)

3. Since different age-sex groups have experienced different lifetime air pollution exposures, we plan to introduce various latency periods for the onset of diseases apparently caused by air pollution. This is a way to conduct sensitivity tests in accordance with the presupposed latency periods.

4. An attempt should be made to define pollutant thresholds in those cases for which some association with mortality had been indicated. If threshold effects can be shown to exist, then the linear model does no longer appear applicable.

5. In order to draw more specific conclusions about age-specific mortality rates, we will partition the age groups into finer brackets, say into time spans of 10 yr.

6. Most people now spend a significant part of their time indoors; thus indoor/outdoor air quality relationships are an important aspect of the estimation of a population's exposure to air pollution.

7. It still remains to be proved whether cigarette smoking and atmospheric pollution do interact. By a modification of the model we could find out whether synergisms exist between smoking and atmospheric pollution as a group, in the same way as synergisms were found between cigarette smoking and asbestos exposure.

Each of these problem areas could be submitted to the model in asking "what if" questions, but naturally constraints exist, in particular with respect to data availability, if one is limited to census-type data.

## Appendix A: Computation of Mortality Rates

These mortality rates were calculated using the following standard demographic procedures (see Spiegelman 1978):

I. 1. List total deaths for all individuals of age group *a*, sex *s*, who resided in census tract *i* at the time of their death during the period 1974–1978.
   2. List deaths from cause *c* in census tract *i* for this age-sex group during these years.
   3. Multiply five times the 1976 population of age *a*, sex *s* in census tract *i*. The reason for this step is that the procedure involves a 5-yr average mortality rate and 1976 is the middle year of this period. It is assumed that the changes in the population during this period occurred at a constant linear rate.

II. Add one-half of item I.1 to the figure in item I.3. The calculation provides an estimate of the population at risk and is based on the assumption that half the deaths precede the midpoint.

III. Divide the figure in item I.2 by the results of step II in order to obtain the desired mortality rate. Algebraically, this calculation may be written as

$$\text{Mortality rate} = \frac{\text{I.2}}{\text{I.3} + [\text{I.1}/2]}.$$

## Appendix B. An Illustrative Model of Balancing Mortality Costs of Pollution by a Cost-Effectiveness Strategy

Suppose that in a number of SMSAs the mortality rating $Y$ is believed to depend on two factors: $SO_2$ air quality $x_1$ and total particulate (TP) air quality $x_2$, measured in some suitable units. The expected mortality rating is assumed to be

$$EY = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2.$$

The policy problem for the pollution control agency is to determine the relative importance of $x_1$ and $x_2$ in the regression function in order to decide where to put the efforts toward decreasing pollution. Thus, for any practical antipollution program a means would be sought to optimize expected mortality ratings for given allocation of funds on $SO_2$ and TP air quality. If $C_1$ and $C_2$ are costs (in money units) of increasing $x_1$ and $x_2$, respectively, by one unit, and if steps taken to improve $x_1$ result in a reduction of $x_2$ by an amount $q_2$, whereas steps taken to improve $x_2$ result in a reduction of $x_1$ by an amount $q_1$, then a program allocating funds $F_1$ and $F_2$ ($F_1 + F_2$ fixed) to improving $x_1$ and $x_2$, respectively, will lead to an expected improvement in morality rating of

$$(F_1/C_1)(\alpha_1 - \alpha_2 q_2) + (F_2/C_2)(\alpha_2 - \alpha_1 q_1).$$

Because of linearity, the most cost-effective program is then to make all funds available to improve $SO_2$ air quality $x_1$, if $(\alpha_1 - \alpha_2 q_2)/C_1 > (\alpha_2 - \alpha_1 q_1)/C_2$ and to improve TP air quality $x_2$ if $(\alpha_1 - \alpha_2 q_2)/C_1 < (\alpha_2 - \alpha_1 q_2)/C_2$.

Note that $q_1$ and $q_2$ may be either positive or negative, depending on circumstances, if these factors are unknown but believed to be small, a satisfactory approximation will be provided by the simplified rule: Improve $x_1$ if $\alpha_1/C_1 > \alpha_2/C_2$, improve $x_2$ otherwise.

The model might include an interaction term, perhaps the product of $x_1$ or $x_2$ with a third variable, and that would make the decision rule more complex. In a more fundamental way, it is possible to extend the analysis to take into account uncertainties in the coefficients and costs. This might lead to a mixed cost-effectiveness strategy, one in which funds are put into improvement of both $SO_2$ and TP air quality.

## References


Auster, R. (1972) The production of health, an exploratory study in *Essays in the Economics of Health and Medical Care,* V. R. Fuchs, ed., pp. 135–158. National Bureau of Economic Reasearch, New York, NY.

Crandall, R. and Lave, L., eds. (1981) *The Scientific Basis of Health and Safety Regulation.* The Brookings Institution, Washington, DC.

Fuchs, V. (1973) Some economic aspects of mortality in developed countries. Presented at the International Economic Association Conference on Economics of Health and Medical Care, Tokyo, April 2-7.

Gottinger, H. W. (1974) Systems approaches and environmental problems in *International Symposium "Schloss Reisensburg."* Göttingen and Zürich: Vandenhoeck & Ruprecht.

Gottinger, H. W., Fritz, A., and Faus-Kessler, M. (1981) Luftverschmutzung and Mortalität (Air pollution and mortality). Durchführbarkeitsstudie im Auftrag des Bayer. Staatministeriums für Landesentwicklung und Umweltfragen München, Federal Republic of Germany.

Harriss, R. C. (1979) The burden of technological hazards, in *Energy Risk Management,* G. T. Goodman and W. D. Rowe, eds. Academic Press, New York, NY.

Katz, M. (1980) Advances in the analysis of air contaminants, *J. Air Pollut. Control Assoc.* **30,** 528–557.

Land, C. E. (1980) Estimating cancer risks from low doses of ionizing radiation, *Science* **209,** 1197–1203.

Landsberg, H. H. (1979) *Energy, The Next Twenty Years,* Chap. 10, 11. Ballinger, Cambridge, MA.

Lave, L. B. and Seskin, E. P. (1972), Air pollution, climate and home heating: Their effects on U.S. mortality rates, *Am. J. Public Health* **62,** 909–916.

Lave, L. B. and Seskin, E. P. (1973) An analysis of the association between U.S. mortality and air pollution, *J. Am. Statist. Assoc.* **68,** 284–290.

Lave, L. B. and Seskin, E. P. (1976) Does air pollution cause mortality? Proceedings of the Fourth Symposium on Statistics and the Environment, American Association for the Advancement of Science: Washington, DC.

Lave, L. B. and Seskin, E. P. (1977) *Air Pollution and Human Health Resources for the Future Study,* Johns Hopkins University Press, Baltimore, MD.

Lave, L. B. and Chappie, M. (1982) Air pollution and human health: A re-examination, *J. Urban Econ.* (in press).

Lipfert, F. W. (1980) Sulfur oxides, particulates, and human mortality: Synopsis of statistical correlations, *J. Air Pollut. Control Assoc.* **30,** 366–371.

McDonald, G. C. and Schwing, R. C. (1973) Instabilities of regression estimates relating air pollution to mortality, *Technomet.* **15,** 463–481.

Ricci, P. F. and Wyzga, R. E. (1980) A statistical review of cross sectional studies of ambient air pollution and mortality, DOE Statistical Symposium, Berkeley, CA.

Riggan, W. B. (1973) Chess, A community health and environmental surveillance system, in *Sixth Berkeley Symposium on Mathematical Statistics and Probability,* J. Neyman, ed., pp. 111–123. University of California Press, Berkeley, CA.

Shy, C. M. (1979) Epidemiologic evidence and the United States Air Quality Standards, *Am. J. Epidem.* **110,** 661–671.

Smith, W. E. (1968) Factors associated with age-specific death rates, *Am. J. Public Health* **58,** 1937–1949.

Spiegelman, M. (1978) Life tables, in *International Encyclopedia of Statistics* W. Kruskal and J. M. Tanur, eds., Vol. 1, pp. 512–518. New York, NY.