MORBIDITY, AIR POLLUTION AND HEALTH STATISTICS

Bart D. Ostro
and
Robert C. Anderson
Economic Analysis Division, PM-220
Office of Policy Analysis
*U.S. Environmental Protection Agency
401 M Street, S. W.
Washington, D.C. 20460
(202) 382-2790

Presented at the Joint Statistical Meetings of the American Statistical Association and Biometric Society
Detroit, Michigan
August 12, 1981

*This paper represents the work and views of the authors. It does not necessarily represent the policy or position of the U.S. Environmental Protection Agency.

The Clean Air Act of 1970 committed the public and private sectors to billions of dollars in expenditures for pollution control. The authority provided by this Act, including the regulation of stationary and mobile sources of pollution, is scheduled for review by Congress this year. Though air pollution is known to affect agricultural output, materials, and visibility, the major thrust of the Act and related regulations issued by the U.S. Environmental Protection Agency is to reduce pollution that causes adverse impacts on human health. The Clean Air Act specifies that national ambient air quality standards should be set to protect the health of sensitive groups with an adequate margin of safety.

A new need to estimate actual health effects, and not just thresholds, for various air pollutants now exists. Executive Order 12291 issued by President Reagan in February 1981 requires consideration of the potential benefits and costs of all major regulations. It is expected that the revision of primary air standards over the next two years will be interpreted as major regulations subject to the requirements of the order.

Recent research has tested empirically the hypothesis that air pollution affects the incidence of human mortality and morbidity. This research has played a major role in the establishment of primary air standards in the United States.

Although potential mortality effects obviously concern policy makers, morbidity effects should be of equal concern to the researcher. First, there may be a good deal of acute illness resulting from air pollution that never results in death. Second, morbidity is probably a more sensitive indicator of

pollution effects because of the immediacy of the effect. Third, it requires a smaller sample size to obtain statistical significance since sickness occurs more frequently than death. Fourth, the measurement of chronic and more severe morbidity can serve as a verification of the estimated mortality effects.

This paper reviews some of the research techniques that have been used to estimate the health effects of air pollution. It highlights some of the reported results and suggests some of the problems inherent in this type of statistical analysis. Finally, it provides some new evidence of the morbidity effects from air pollution based on some recently developed data.

STATISTICAL TECHNIQUES

Four principal approaches have been used to assess the impact of air pollution on human health. The studies may be broadly characterized as those involving animals and those involving humans. Extensive use has been made of animal studies, especially with respect to mechanisms of damage to pulmonary macrophages. Animal studies have the great virtues of permitting measurement of long-term responses to relatively low levels of pollution and of permitting direct examination of damaged tissue. However, they have the the fundamental problem of extrapolation to man. With this link being imprecise at best, animal studies can play only a minor role in the estimation of pollution control benefits.

Researchers have conducted at least three types of human studies - chamber experiments, statistical analyses of occupational exposure, and epidemiologic studies of general populations.

Although chamber studies on healthy and diseased subjects can reveal the levels at which various acute effects are observed in humans, they have other problems such as the ethics of research and the difficulty of ascertaining chronic effects. Chamber studies can have great value in identifying thresholds for various subgroups of the population, but they can play only a limited role in assessing economic benefits because of the virtual omission of consideration of chronic effects. They typically measure changes in metabolism or organ function, as opposed to changes in human activity. Further, because of the heterogeneous mixture of suspended particles found in the environment, comparisons between chamber studies of laboratory-produced particles and actual human exposure is extremely difficult.

Studies of occupational exposures face problems of measuring actual exposures (particularly a problem for chronic effects resulting from past exposure) and the typical lack of compatability between the mix of occupational pollutants and concentrations and those experienced by general populations. Moreover, there is substantial evidence of selection by industry and self selection by workers so that exposured industrial groups are not representative of the population. Therefore, occupational data will have limited relevance to the estimation of benefits of control of population exposures to the air pollutants of concern to policy makers.

Because of the problems with the aforementioned approaches, assessments of general population benefits from controlling air pollution must rely on epidemiologic methods in general populations to detect chronic effects of long-term exposures to low

levels of pollution. The epidemiologic approach has the great virtue of being able to estimate the response to the full range of conditions to which humans are actually exposed. Its main drawbacks are the difficulty of controlling all potentially confounding variables and it can at best establish only correlation and not causality. Microepidemiologic studies using data on individuals are preferable to studies that use data on the means of large groups exposed to different levels of pollution because data on individuals spans a far larger range of variation than do data on, say, metropolitan area averages. For example, cigarette smoking is a major contributor to adverse health states, probably much more important than than air pollution parameters. The use of citywide averages may obscure statistically significant dose-response relationships that exist among the nonsmoking subset of the population. Despite the preferability of data on individuals, such data are often unavailable, and many researchers have usually relied upon average responses and average characteristics for population subgroups as the basis for analysis.

Statutory obligations (in the Clean Air Act of 1970) forced regulators to set air pollution standards to provide a margin of safety in protecting sensitive groups. This statutory directive undoubtedly shifted research interests toward defining threshold levels for sensitive groups. For this purpose, chamber studies were indispensable. Interest in population epidemiologic approaches, however, continued principally because it was the only method for possibly evaluating the incremental impacts on the general population of alternative air standards.

RESULTS OF PREVIOUS STUDIES

In the 1960s and into the 1970s, several epidemiologic studies and chamber experiments were reported. They concerned the health consequences of both short-term and chronic exposures to particulates and sulfur oxides (including sulfate).* A number of the earlier epidemiologic studies of particulates and SO_2 focused on severe air pollution episodes such as those in London, New York City, and Donora, Pennsylvania. In London, 4,000 excess deaths were attributed to one severe episode in 1952. Morbidity effects from acute exposures, measured in terms of hospital emergency room visits, doctor visits, and industrial sickness records also have been demonstrated.

Long-term health effects of chronic exposures, have also been documented. In this brief summary of this work, we will focus on two types of studies: epidemiological studies of mortality and morbidity and approaches attempting to measure the increased risk of long-term exposure.

Beginning with the pioneering work of Lave and Seskin, a number of studies have used regression analysis to measure the long-term effects of sulfur oxides and particulates on mortality.1/ Lave and Seskin attempted to explain several

*We have concentrated on SO_2, SO_4, and particulates in this analysis because of the relative wealth of epidemiologic evidence linking them to chronic health effects in humans (as contrasted to CO, NO_X, or ozone).

1/Lave, Lester, and Eugene Seskin, "Air Pollution and Human Health," Science Vol. 169, August 21, 1970, pp. 722-733, and Lave and Seskin, Air Pollution and Human Health (Baltimore: Johns Hopkins University Press, 1977.)

disease-specific mortality rates using cross-sectional data from 177 standard metropolitian statistical areas. Their work has been viewed with some degree of caution because of such concerns as: (1) the estimation bias resulting from omitted explanatory variables; (2) the omission of personal factors, such as age, sex, and cigarette consumption; (3) failure to control for in- and out-migration; (4) crude measurement of exposure, and (5) failure to fully consider alternative functional forms. Despite these shortcomings, the basic results suggesting an association between air pollution and mortality have held up over time under careful scrutiny by subsequent researchers.

Further studies by others have attempted, with varying degrees of success, to correct one or more of these deficiencies. Mendelsohn and Orcutt[2/] controlled for migration, Crocker et al.,[3/] for medical inputs and diet, and Lipfert,[4/] for other socioeconomic variables. Crocker et al. and Gregor[5/] also corrected for possible simultaneity with physician services. These

---

2/Mendelsohn, Robert, and Guy Orcutt, "An Empirical Analysis of Air Pollution Dose-Response Curves," Journal of Environmental Economics and Management, Vol. 6, June, 1979.

3/Crocker, T.D., W.D. Schulze, S. Ben-David, and A.V. Kreese, Methods Development for Assessing Air Pollution Control Benefits, U.S. Environmental Protection Agency, February, 1979.

4/Lipfert, Frederick W., "Statistical Studies of Mortality and Air Pollution: Multiple Regression Analyses Stratified by Age Group," mimeo, 1979.

5/Gregor, John J., Intra-Urban Mortality and Air Quality: An Economic Analysis of the Costs of Pollution Induced Mortality, Environmental Protection Agency, Corvallis, Oregon, 1977.

studies have obtained a range of estimates of .01 to about .2 for the elasticity of mortality with respect to air pollution. As a result, Freeman[6] in his synthesis of the literature, chose .05 as the best point estimate of the elasticity. There is still much concern, however, about the legitimacy of these studies as evidence of the chronic effects of air pollution.

Fewer studies have used multivariate techniques to estimate morbidity effects for air pollutants. A few that should be noted include Crocker et al., who used the Michigan Survey Research Center interview data; Graves and Krumm,[7] who examined data on emergency room visits in Cook County, Illinois; Seskin,[8] who studied unscheduled visits to health clinics in Washington, D.C.; and Liu and Yu,[9] who used a novel two-stage approach to deal with multicollinearity. These studies have typically found associations between particulates and/or sulfur oxides and morbidity measures. Many of theses studies, however, have suffered from serious methodological shortcomings or data deficiencies.

---

[6] Freeman, A. Myrick, "The Benefits of Air and Water Pollution Control: A Review and Synthesis of Recent Estimates," for The Council on Environmental Quality, 1979.

[7] Graves, Philip E., and Ronald J. Krumm, "Pollution and Hospital Admissions: Evidence from Time Series in Chicago, ERC Research Report, 78-1a, 1979.

[8] Seskin, Eugene P., "An Analysis of Some Short Term Health Effects of Air Pollution in the Washington, D.C. Metropolitan Area," Journal of Urban Economics, Vol. 63, July 1979.

[9] Liu, Ben-Chieh, and Eden S. Yu, Physical and Economic Damage Functions for Air Pollutants by Receptor, U.S. Environmental Protection Agency, Corvallis, Oregon, 1976.

Because of the difficulties of measuring actual exposures and also of controlling for diet, smoking, and other personal characteristics in the epidemiological approach, a number of researchers have resorted to expensive case control studies in which individuals are monitored over relatively long periods for effects such as cough, sputum, and respiratory disease. These studies have shown demonstrable correlations between the frequency of symptoms or diseases of the respiratory tract and air pollution levels.

For example, Lunn et al.[10] found a significant relationship between respiratory illness and air pollution among children living in different parts of Sheffield, England. Rudnik[11] documented a relationship between respiratory illness and more polluted cities in Poland. In studies of adults, both Ferris[12] and Bouhuys et al.[13] recorded an assoication between higher levels of TSP and increases in the rates of respiratory disease symptoms.

Epidemiologic evidence has played a central role in the establishment of primary air standards in the United States.

10/Lunn, J.E., J. Knowelden, and A.J. Handyside, "Patterns of Respiratory Illness in Sheffield Infant Schoolchildren," British Journal Prev. Soc. Med., Vol. 21, 1967.

11/Rudnik, J., "Epidemiological Study on Long-Term Effects on Health of Air Pollution," Probl. Med. Wieku Rozwojowego, Vol. 7a (Suppl.1), 1978.

12/Ferris, B.G. Jr., I.T.T. Higgins, M.W. Higgins and J.M. Peters, "Chronic Non-specific Respiratory Disease in Berlin, New Hampshire, 1961-1967. A follow-up study," Am. Rev. Resp. Dis., Vol. 107, 1973.

13/Bouhuys, A., G.J. Beck, and J.B. Schoenberg, "Do Present Levels of Air Pollution Outdoors Affect Respiratory Health," Nature, Vol. 276, 1978.

Because of the continuing controversy over chronic morbidity effects at exposure levels near or below the present U.S. standard, better data and improved model specification will be necessary if epidemiologic evidence is to resolve the issue. In this spirit, we have obtained access to much better data than has heretofore been analyzed.

PROBLEMS WITH ESTIMATING HEALTH EFFECTS

Some of the statistical problems of an epidemiological approach to estimating the morbidity effects of air pollution are common to almost all areas of statistical inquiry. Others are more specifically related to uncertainty in the measurement of air pollution and health. The statistical problems can be generalized into three different areas: questions of proper functional form, data and measurement problems, and specification problems and uncertainties.

Functional Forms

Most epidemiological research on the health effects of air pollution has assumed a linear dose-response relationship. The additive linear functional form implies that each marginal improvement in air quality results is a constant improvement in health. In addition, it posits that there are no interactive effects among pollutants or between pollution and other variables, such as weather conditions.

Unfortunately, there is little theoretical or empirical justification for this functional form. Most clinical research has generated an S-shaped (or logistical) dose-response relationship.

However, the assumption of linearity may be an acceptable approximation of the true form over a certain range of air pollution values. For large changes in air pollution, the linear approximation will likely be a less accurate estimate of the health effects than some nonlinear specification.

There are two other potential problems with the linear form. First, it can predict negative values for the dependent variable, even if the dependent variable is always observed to be non-negative. Second, it structurally assumes that the explanatory variables will have a similar effect over the entire range of the dependent variable.

If one is attempting to estimate the probability of death or illness from air pollution and wishes to use a nonlinear functional form, a number of probabilistic models are available, including logit, probit, and Tobit. Each carries its own assumptions about the shape of the dose-response function and about the error term. With the uncertainty intrinsic to an area of inquiry such as air pollution and health, it is extremely important that alternative function forms be tested to compare the goodness-of-fit, compatability, and predictive results.

Data and Measurement Problems

The second major statistical problem germane to the study of air pollution and health is that of availability and accurate measurement of the necessary data, especially of air pollution exposure. There are three major concerns here: which pollutant to measure, the relationship of ambient levels to actual exposure, and the time structure of pollutant exposure.

Measurements of ambient air pollution are obtained primarily through Environmental Protection Agency monitors sited throughout the country. The measurement techniques have improved dramatically over the last two decades and are becoming more accurate and specific. For example, EPA is moving towards measuring and setting standards for inhalable particulates (those less than 15 microns) which are now believed to be more harmful to the respiratory system than total suspended particulates. Among the other pollutants, however, there is still question as to which are the most important precursors of health effects. Only further clinical study will reduce uncertainty in this area.

Even the most accurate measurement of pollution at the monitoring site may not represent the measure of actual pollution exposure, however. First, there can be significant spatial variation of the pollutant and the potential receptor around the source of measurement. Second, individuals working in other areas or in closed environments will receive different exposures for at least part of the day. Third, actual exposure will vary according to the time spent inside, the degree of insulation and ventilation, and the prevalent pollutant in the area. For example, carbon monoxide easily penetrates all structures, while large-order particulates and reactive pollutants, such as sulfur dioxide and ozone, do not. Researchers usually make the simplifying assumption that, on average, the monitored air quality level is somewhat representative of exposure. Random measurement errors of air pollution exposure should lead to an estimate of the air pollution effect that is biased towards zero.

Another question relating to the use of ambient levels as a proxy for exposure is that of which statistical measure to use. The mean, maximum, and minimum pollution levels all have been used in the past. Each suggests a different kind of relationship between air pollution and health. Satisfactory answers to this question would help the policy makers decide if it is chronic doses above some minimum level or acute doses at high levels that generate serious health effects.

Finally, there is a question of the time lag of health effects caused by air pollution. Health effects may well be related to current levels of pollution, or they may be a result of cumulative exposure over a number of years. If the latter is the case, the use of current levels may lead to a biased estimate of the pollution effect.[14]/

The choice of the health measure also presents a problem for morbidity research. Although there are many of sources of data on illness and hospital visits, few have the standardization and sample size necessary for a cross-sectional analysis. Thus, most of the morbidity studies have been either time series analyses for a given city, studies of emergency room utilization, or simple two-city or city-rural comparisons using analysis of variance. In addition, some surveys which have attempted to link overt effects - e.g., eye stinging, sneezing, coughing, and breathing - with

---

[14]/See Daniel M. Violette, "Estimating the Human Health Benefits of Improved Air Quality," prepared for the National Commission on Air Quality Benefits Estimation Panel, January 1980, pp. 189-193.

recorded levels of air pollution. Recently, some other data bases, which include questions about health care utilization and health status, have been used. These include the Michigan Survey Panel Data and the National Center for Health Statistics Health Interview Survey (HIS).

The HIS has many possibly useful indicators of health status. For acute illness, it measures restricted activity days, work loss days, school loss days, bed days, and hospital days. Restricted activity days (RAD) is the inclusive term for all the ways one can react to acute illness. It is officially defined in the HIS as a day in which "a person cuts down on his usual activities for the whole of that day because of an illness or injury.... It does not imply complete inactivity, but it does imply only the minimum of usual activities." In addition, the HIS reports the health condition or diagnosis that is believed responsible for each RAD.

The variable measuring work loss days is based on the response to the survey question asking how many days in the last two weeks did illness or injury prevent one from working. Obviously, the amount of pain or discomfort tolerated by an individual before missing work is a very subjective decision and may have little to do with any objective measure of illness. In addition, reported or actual WLDs may be affected by other unmeasured factors, such as response to the survey or attitude toward work. Part of the decision to miss work, however, will be based on socioeconomic and job-related factors that can be measured or approximately empirically. The statistician can only assume that there is an underlying distribution that determines the threshold

of health effects. For each chronic illness, the HIS records the duration of limitation, the degree of limitation and the diagnosis.

The measurement of other, potentially confounding variables is also important to the study of air pollution and morbidity, especially since the "true" causative variables to describe morbidity are unknown. Omission of variables that explain the variation in the dependent variable can lead to serious estimation problems.

Much of the previous research on health effects has used aggregate data to proxy socioeconomic variables. For example, in their mortality study, Lave and Seskin use such variables as the percentage of population 65 or older, the percentage of the population who are nonwhite, and the percentage with income below the poverty level. Individual data represent a distinct improvement and allow the researcher to disaggregate the analysis and discern the variation in the pollution effect across categories, such as age, race, and sex.

A number of other variables may vary collinearily with air pollution and may also affect health status. Those most frequently cited factors include occupational exposure, smoking, migration, indoor pollution, diet, exercise, risk attitude, weather, and "urbanness."15/ Again, some of these confounding

---

15/For a detailed description of the problems generated by these factors and attempts to reconcile them in mortality studies see Richard Wilson, et al., Health Effects of Fossil Fuel Burning, (Cambridge: Ballinger Press, 1980), pp. 191-214.

effects can be eliminated through using individual data, if available. By stratifying the sample one can explicitly account for the effects of occupation, smoking, indoor exposure, and various geographic factors.

For some factors, such as diet, exercise, and attitude towards health care, direct measurement through survey will probably not be economically feasible. However, many of these influences can be proxied by socioeconomic surrogates. A statistically significant pollution effect can be generated artificially only if these factors vary with air pollution and not with the socioeconomic proxies.

Specification Problems

Even if acceptable data on pollution exposure, health status, and their potentially confounding factors are available, improper specification of an estimated equation can seriously bias the coefficients. Three different specification problems may be relevant to this area of research: multicollinearity, omitted variables, and simultaneity.

Since the "true" model of health status is far from certain, one can only make reasonable guesses about the variables that should be included in a regression equation explaining illness. A trade-off is involved. As explanatory variables are added, multicollinearity may became a problem; specifically, variables that vary with air pollution may be included so that the estimated effect of pollution becomes confounded. To limit the number of explanatory values, however, is to open up the possibility of omitted variable bias.

Multicollinearity can exist, and usually does, among air pollution variables. Particulates, sulfur dioxide and sulfates are all generated from fossil fuel combustion by stationary sources. On the other hand, hydrocarbons, carbon monoxide, nitrous oxides, and ozone are primarily the result of fuel combustion from mobile sources. Multicollinearity can also arise because of the relationship between air pollution and the other explanatory variables including socioeconomic and urbanization variables. To the extent that these factors vary systemically (e.g., both air pollution and urbanization may increase as we move from the southwestern to the northeastern United States), discerning the independent influences of air pollution will be difficult.

Another potentially serious specification error occurs when a nonrandom explanatory variable, correlated with air pollution, is omitted from the estimated equation. The included independent variables then take on explanatory "noise" from both the excluded variable and the error term and will have biased estimated coefficients. The degree of the bias will be proportional to (1) the collinearity between the excluded and air pollution variables, and (2) the importance of the omitted variable in explaining the dependent variable.

A final specification problem is that of simultaneity. This would occur if, for example, the explanatory variable "physicians per capita" is used to explain the variation in health status. If health status in turn influences the locational decision of physicians, the estimated coefficients will be biased and inconsistent.

A technique, such as two-stage least squares, could be used to used to reduce this problem.

PRELIMINARY RESULTS

The data set that comes closest to meeting many of the needs outlined above is the annual Health Interview Survey (HIS) conducted by the National Center for Health Statistics. This is a scientific survey of 50,000 households comprising roughly 120,000 people. Besides basic demographic and economic characteristics of the respondents, the survey includes data on acute and chronic illness (identified by diagnosis), disability days for those in and out of the labor force, work and school loss days due to illness, measures of health care utilization, height and weight, family income, occupation and industry of employment, and individual cigarette consumption. The availability of the latter makes the data set superior to many others and facilitates the separation of health effects from cigarette smoking versus air pollution.

For a preliminary assessment of the effects of air pollution on morbidity, a data base was created that provides detailed information about the individuals and their health status, the levels of several pollutants to which they are exposed, their climate, and the area where they live. Thus, the HIS results for 1976 were merged with 1976 EPA data on ambient levels of particulates (TSP), sulfur dioxide (SO_2), and sulfates (SO_4); National Oceanic and Atmospheric Administration data on wind, temperature, and precipitation; and Census Bureau data on density and other

urban characteristics. For this analysis, 120 cities, most of medium size (population of 100,000-600,000), were preselected to reduce the intracity variation of the air pollution measures.

The initial work focused on determining the contribution of air pollution to acute illness in adults. The sample of all male nonsmokers was used to estimate the variation in work loss days (WLD). This group was chosen for a number of reasons. First, the sample size of males is greater than that of females. Second, with nonsmokers the air pollution effects cannot be attributed to the impact of cigarette smoking.* Also, cigarette smoking may be determined simultaneously by variables that are used to explain health status. If smoking were included as an explanatory variable, it would necessitate a slightly more complex and less easily interpreted model. Third, males tend to have less family and child-rearing responsibilities outside of work. Therefore, there is less of a possibility of the occurrence of work loss days not related to health. Thus, work lossd may be a more accurate indicator of illness for males than for females. Finally, measuring work loss is mroe conducive to a monetary evaluation of losses.

The dependent variable was hypothesized to be a function of levels of ambient air pollution, various demographic and socioeconomic variables, the exitence of chronic disease, climate conditions, and measures of "urbanness."

---

*There still remains the possibility that nonsmokers living with a smoker will be affected by the smoke. This possibility will be considered in subsequent work.

Basically, two pollution variables, total suspended particulates TSP and sulfates, were used. They were selected because of the preponderance of clinical evidence previously mentioned concerning their health effects and because their measurement tends to be acceptably consistent. The correlation coefficient of these two variables was .18.

In the past, concern has been expressed about the choice of the measure of pollution exposure for a city. For the purposes of this study, the SAROAD system, EPA's aerometric data bank was used. For many cities in the sample, there was only one population-oriented monitor. For cities with more than one population-oriented monitor, a weighted average of the monitors, based on the number of observations, was calculated. The TSP and sulfate measurements were based on recordings from hi-vol 24-hr gravimetric samplers and hi-vol colorimetric samplers, respectively.

A number of demographic and socioeconomic variables -- including age, race, family income, family size, physicians per 100,000 people, blue- or white-collar worker, and whether or not the individual was married and currently living with spouse -- were all employed to explain the variation in WLD. These variables were believed to be important factors in measuring the degree of and response to pollution exposure and the ability to partake in preventive care, including direct physician access, housing and sanitary conditions, diet, exercise, and occupational exposure. Data limitations preclude a determination of the degree to which diet and exercise, for example, may affect health.

It is believed however, that the included independent variables are ample proxies for the measurement of access to, and use of, preventive care while at the same time independent enough to preclude problems with multicollinearity.

The existence of chronic disease (a binary variable) will probably play an important role in determining the frequency of work loss or activity restriction and was included in the estimation. The climatic conditions faced by individuals, such as precipitation and average temperature or number of degree days, were considered because of their potential effect on WLD. Finally, population density was included as a measure of the general urban structure.

Multiple regression was selected as the appropriate statistical tool because of its ability to control for many factors in the analysis. A major uncertainty in the estimation, however, was the exact form that the statistical model should take. A special problem exists in that the dependent variable is truncated at zero, and that a large percentage of the health status observations (between 70 and 95 percent) are zero. For this reason, three different models were tested. Each has different characteristics and assumptions about the structural nature of the explanatory variables, and each generates a different shape for the dose-response relationship.

First, the ordinary least squares (OLS) method was used. Although cheaper to run and computationally simpler, this technique ignores the zero truncation and can possibly predict negative values for WLD. In addition, it has the implicit structural

assumption that the same factors that cause the existence of any work loss day (the movement from zero to one or more) also explain the particular number of WLDs, given that at least one WLD has actually occurred. One advantage to this technique is that linearity makes extrapolation easier. The estimated equation took the following form:

$$(1) \quad W_1 = b_o + b_1 D + b_2 A + b_3 C + b_4 M + b_5 U + u$$
$$= bX + u$$

where W_1 = Number of work loss days

b_i = Estimated coefficients

D = Demographic and socioeconomic characteristics

A = Air pollution measures

C = Chronic condition

M = Meteorologic variables

U = Urban structure variables

b = Vector of the coefficients

X = Vector of the above independent variable

The partial derivation of work loss days with respect to the air pollution variable is:

$$(2) \quad \partial W_1 / \partial A = b_2$$

An alternative technique was to use the Tobit model. This technique constrains the dependent variable to be non-negative but still implies the structural assumption described above. An additional problem is that the shape of the resulting dose-response curve will have positive first and second derivatives

(convex from below), which is contrary to the generally accepted shape of the curve.

The stochastic model underlying the Tobit estimation is:

$$(3)\quad W_2 = bX + u \quad \text{if } bX + u > 0$$

$$W_2 = 0 \quad \text{if } bX + u \leq 0$$

$$\text{with } u \sim N(0, \sigma^2)$$

where W_2 = proportion of all work days that are lost days. The model assumes that there is an underlying stochastic index $I = Xb + u$ that is observed only when it is positive.

Following Tobin,[16] the expected value of W_2 in the model is:

$$(4)\quad EW_2 = bX\,F(Z) + \sigma f(Z)$$

where $Z = bX/\sigma$, $f(Z)$ is the unit normal density, and $F(Z)$ is the probability that a normalized random variable with a zero expected value will take a value less than or equal to Z.

The expected value of W_2 for observations above the limit, W_2^*, is expressed by:

$$(5)\quad EW_2^* = bX + \sigma f(Z)/F(Z)$$

The latter term is the expected value of the truncated normal error term. The meaning of F(Z) is simply the probability of the value Z being above zero.

16/Tobin, James, "Estimation of Relationships for Limited Dependent Variables," Econometrica, Vol. 26, January 1958.

Following McDonald and Moffitt,[17] the relationship between the expected value of all the observations, W_2, the expected value of those values above zero, W_2^*, and F(Z) is:

$$(6) \quad EW_2 = F(Z)\, EW_2^*$$

The partial derivative of the expected value of all observations expressed in (6) with respect to air pollution is:

$$(7) \quad \partial EW_2/\partial A = F(Z)(\partial EW_2^*/\partial A) + EW_2^* \, (\partial F(Z)/\partial A)$$

or the change in W_2 for those observations above zero weighted by the probability of being above the limit plus the change in the probability of being above zero weighted by the expected value of of W_2 if above zero. With estimates of b and σ, both of the terms on the right-hand side can be calculated.

The third technique used was the logit-linear model. In this case, a logit model was first used to determine the probability of a person's having at least one WLD in the survey period. In the second stage, OLS is used to determine whether air pollution influences the number of WLDs, given a person's has had at least one. This method has the advantage of consistency with statistical characteristics of the data. First, it truncates the dependent variable at zero (and one) by turning the frequency into a probability. Second it enables the use of different structural forms to explain the probability of a WLD episode (one or more) and the number of

---

[17] McDonald, John F., and Robert A. Moffitt, "The Uses of Tobit Analysis," Review of Economics and Statistics, Vol. 62, No. 2, May 1980.

WLDs. Third, the estimated equation will assume the form of the logistic curve, the functional form that is believed to be typical of many dose-response relationships.

The estimated equation of the logit model is:

$$(8) \quad \log [(W_3/(1 - W_3)] = Xb$$

where W_3 is the probability that WLD > 0 in the two-week survey period. The left-hand side of (8) is simply the log of the odds of a work loss day. The change in W_3 due to a change in A is:

$$(9) \quad \partial W_3/\partial A = b_2 \cdot W_3 (1 - W_3)$$

The equation can also be expressed in terms of probability:

$$(10) \quad W_3 = (1 + e^{-Xb})^{-1}$$

The expected number of work loss days is the product of the probability of a nonzero WLD times the number of WLDs,

$$(11) \quad E(W) = W_3 \cdot E(W_1 \mid W_1>o)$$

The regression results for the three models using the sample of all male nonsmokers, age 18-65, are presented in Table 1. The results of the three estimates are generally consistent with prior expectations. In all three models, particulates are shown to be related in a positive and significant way to work loss days.*

*The value of the particulate variable ranges from 43 to 150. Subsequent analysis suggested that the particulate coefficient was statistically significant from zero when TSP was as low as 65 to 70 micrograms. The current annual standard is 75 micrograms.

The mean level of sulfates does not appear to affect WLDs. This result was confirmed when each of the pollution variables was run separately in the regression.

There may be a number of explanations for this result. First, the techniques for measuring sulfates are not believed to be very accurate. The errors in measurement may lead to serious underestimation of the coefficient. Second, the particulate measure may be proxying a number of variables; it measures coarse and inhalable particles as well as sulfate and nitrate particles. Third, there may be estimation problems resulting from collinear or omitted variables.

Using equations, (2), (7) and (9) the partial affects of air pollution can be calculated. The results indicate that the OLS, Tobit and logit, models predict that a one unit change in TSP will, at the mean, change the probability of a work loss day in the two week period by .00177, .00118, and .0013 respectively. For the OLS model, the work loss-particulate elasticity, measured at the mean, is 0.57.

The models also show that chronic illness is associated with more WLDs. The OLS model has age and average temperature related positively to WLD and blue-collar employment related negatively to WLD. Comparing this model to the logit-OLS model some interesting distinctions can be made.

Estimation (C) suggests that air pollution, measured by particulate levels, will affect the probability of a WLD episode. However, estimation (D) suggests that air pollution does not influence the number of days lost, given an episode has occurred.

TABLE 1

The Estimation for WLD for Male NonSmokers (N=4473)

	(A) OLS	(B) TOBIT	(C) LOGIT	(D) OLS(W>1)
CONSTANT	-.29	-1.2	-3.97	-.61
PMEN	.00177[b]	.00228[b]	.00614[b]	.0032
SULF	-.0083	-.013	-.033	-.0055
AGE	.004[a]	-.0003	-.0076[c]	.0746[a]
CHRON	.22[a]	.16[a]	.35[a]	1.25
RACE	-.02	-.027	-.04	-.36
MARR	-.03	-.0012	.37[a]	-.9[a]
INC	-.0045	-.0014	-.0023	.0041
TEMP	.0063[b]	.0014	.005	.088[a]
PRECIP	-.0013	.0024	.009	-.04[c]
DENS	.004	.0069	.003	-.027
BLUE	-.06[c]	-.017	.157[c]	-1.33[a]
F	4.07[a]	--	--	6.38[a]
χ^2	--	113.2[a]	25.8[a]	--
R^2	.01	--	--	.19

a = Significance at 1% level
b = Significance at 5% level
c = Significance at 10% level

PMEN = annual arithmetic mean of particulates (micrograms/cubic meter)

SULF = annual arithmetic mean of sulfates (micrograms/cubic meter)

AGE = age

CHRON = number of chronic conditions

RACE = 1 if nonwhite
0 if white

MARR = 1 if married and living with spouse
0 if unmarried or married and not living with spouse

INC = family income (thousands)

TEMP = annual mean temperature

PRECIP = annual precipitation

DENS = population density (thousands)

BLUE = 1 if blue-collar worker
0 if not

Further evidence of this result is obtained by applying Eq(7) to the Tobit estimates. The result, after taking the partial derivative, indicates that the first term in the right-hand side of Eq(7) -- the change in WLD for those observations above zero -- is small (.0000175) relative to the second term -- the change in the probability of being above zero (.0001). Thus, the total effect of air pollution on WLD is driven more by adding to the probability of an episode than by affecting the actual number of WLDs.

The estimated equations (C) and (D) also show that being married and working in a blue-collar job increase the probability of a work loss episode but have a negative effect on the number of days lost. Age has the reverse affect: it slightly decreases the probability of an episode but has a strong positive influence on the number of days lost. The latter result is confirmed by the linear estimate (A).

The sensitivity of the variables was further tested by considering various other subsamples and specifications. For example, the model was estimated for those aged 45-65 and for those with chronic conditions. In each, the magnitude of the estimated air pollution coefficient increased and remained significant.

In addition, other weather and urban variables were substituted into the regresssion with no appreciable change in the estimates. The results of these statistical tests appear to confirm the hypothesized association between air pollution and morbidity.

CONCLUSION

The pending revision of U.S. primary air standards and the analytic requirements of Executive Order 12291 will force regulators to examine closely the data showing possible human health effects from air pollution. Four principal approaches have been used to assess these health effects: animal experiments, chamber studies on human, statistical analyses of occupational exposures, and epidemiologic studies in general populations. Of the four techniques, the epidemiologic approach has the great virtues of including the full range of exposures to air pollution, the various combinations of air pollutants to which humans are exposed, and other possibly synergistic and antagonistic parameters such as smoking and medical care. Of course, the very complexity of these interactions necessitates that great care be given to model specification and the inclusion of all relevant factors.

A number of studies have investigated the relationship between air pollution and human morbidity and mortality using the epidemiologic approach. Sulfur oxides and particulates have been linked to both morbidity and mortality effects in studies of pollution episodes as well as from long-term exposure to lower levels of pollution. Critics have identified several shortcoming that plague many of these studies including (1) omitted variables such as diet and cigarette consumption, (2) poor control for migration, (3) crude measurement of exposure, (4) failure to fully consider alternative functional forms and possible simultaneous relationships, and (5) use of city average data rather than data on individuals within cities.

This study uses a data set on individuals, the Health Interview Survey, conducted by the National Center for Health Statistics to examine further the relationship between air pollution and various measures of morbidity. The scientific survey of 50,000 households in the HIS includes data on demographic characteristics, acute and chronic illness, disability days for those in and out of the labor force, work and school loss days due to illness, measures of health care utilization, family income, occupation, and cigarette smoking. This data was merged with EPA data on ambient levels of particulates, sulfur dioxide, and sulfates; NOAA weather data; and Census Burea data on density and other urban characteristics.

Three regression specifications were used: logit, Tobit, and ordinary least squares. The resulting estimates were generally consistent with prior expections. In all three models, using a sample of male nonsmokers, particulates were shown to be related in a positive and significant way to work loss days. Various tests of the sensitivity of the results using subsamples of the data and alternative specifications all appear to confirm the hypothesized association between air pollution and morbidity.