EPA-230-12-85-020
September 1985

METHODS DEVELOPMENT FOR ENVIRONMENTAL CONTROL BENEFITS ASSESSMENT

Volume II

SIX STUDIES OF HEALTH BENEFITS FROM AIR POLLUTION CONTROL

by

Scott E. Atkinson, Thomas D. Crocker, Ralph C. d'Arge
Shelby Gerking and William D. Schulze
University of Wyoming
Laramie, Wyoming 82701

Shaul Ben David and Reza Pazand
University of New Mexico
Albuquerque, New Mexico 87131

Curt Anderson
University of Minnesota
Duluth, Minnesota 55812

Robert Buechley
Santa Rosa, California 95405

Maureen Cropper
University of Maryland
College Park, Maryland 20742

Larry S. Eubanks
University of Oklahoma
Norman, Oklahoma 73019

Lawrence A. Thibodeau
Educational Testing Service
Princeton, New Jersey 08541

USEPA Grant #R805059-01-0

Project Officer

Dr. Alan Carlin
Office of Policy Analysis
Office of Policy, Planning and Evaluation
U.S. Environmental Protection Agency
Washington, D.C. 20460

OFFICE OF POLICY ANALYSIS
OFFICE OF POLICY, PLANNING AND EVALUATION
U.S. ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

OTHER VOLUMES IN THIS SERIES

Volume 1, Measuring the Benefits of Clean Air and Water, EPA-230-12-85-019.

This volume is a nontechnical report summarizing recent research for EPA on methods development for better estimates of economic-benefits from environmental improvement. The report.,presents the basic economic concepts and research methods underlying benefits estimation as well as a number of case studies, including several from other volumes of this series. Finally, it offers insights regarding the quantitative benefits of environmental improvement.

Volume 3, Five Studies on Non-Market Valuation Techniques, EPA-230-12-85-021.

This volume presents analytical and empirical comparisons of alternative techniques for the valuation of non-market goods. The methodological base of the survey approach - directly asking individuals to reveal their preference in a structured hypothetical market - is examined for bias, replication, and validation characteristics.

Volume 4, Measuring the Benefits of Air Quality Changes in the San Francisco Bay Area: Property Value and Continent Valuation Studies, EPA-230-12-85-022.

This volume replicates a property value study conducted in the Los Angeles Basin for the San Francisco Bay area. A taxonomy series of air quality types and socioeconomic typoligies are defined for cities in the area to examine how property values vary with pollution levels. The contingent valuation method surveys individuals, directly asking their willingness to pay for changes in air quality. The survey method yields benefit values that are about half the property value benefits in both the Bay area and Los Angeles.

Volume 5, Measuring Household Soiling Damages from Suspended Particulate: A Methodological Inquiry, EPA 230-12-85-023.

This volume estimates the benefits of reducing particulate matter levels by examining the reduced costs of household cleaning. The analysis considers the reduced frequency of cleaning for households that clean themselves or hire a cleaning service. These estimates were compared with willingness to pay estimates for total elimination of air pollutants in several U.S. cities. The report concludes that the willingness-to-pay approach to estimate particulate-related household soiling damages is not feasible.

Volume 6, The Value of Air Pollution Damages to Agricultural Activities in Southern California, EPA-230-12-85-024.

This volume contains three papers that address the economic implications of air pollution-induced output, input pricing, cropping, and location pattern adjustments for Southern California agriculture. The first paper estimates the economic losses to fourteen highly valued vegetable and field crcps due to pollution. The second estimates earnings losses to field workers exposed to oxidants. The last uses an econometric model to measure the reduction of economic surpluses in Southern California due to oxidants.

Volume 7, Methods Development for Assessing Acid Deposition Control Benefits, EPA-230-12-85-025 .

This volume suggests types of natural science research that would be most useful to the economist faced with the task of assessing the economic benefits of controlling acid precipitation. Part of the report is devoted to development of a resource allocation process framework for explaining the behavior of ecosystems that can be integrated into a benefit/cost analysis, addressing diversity and stability.

Volume 8, The Benefits of Preserving Visibility in the National Parklands of the Southwest, EPA-230-12-85-026.

This volume examines the willingness-to-pay responses of individuals surveyed in several U.S. cities for visibility improvements or preservation in several National Parks. The respondents were asked to state their willlingness to pay in the form of higher utility bills to prevent visibility deterioration. The sampled responses were extrapolated to the entire U.S. to estimate the national benefits of visibility preservation.

Volume 9, Evaluation of Decision Models for Environmental Management, EPA-230-12-85-027.

This volume discusses how EPA can use decision models to achieve the proper role of the government in a market economy. The report recommends three models useful for environmental management with a focus on those that allow for a consideration of all tradeoffs.

Volume 10, Executive Summary, EPA-230-12-85-028.

This volume summarizes the methodological and empirical findings of the series. The concensus of the empirical reports is the benefits of air pollution control appear to be sufficient to warrant current ambient air quality standards. The report indicates the greatest proportion of benefits frcm control resides, not in health benefits, but in aesthetic improvements, maintenance of the ecosystem for recreation, and the reduction of danages to artifacts and materials.

DISCLAIMER

This report has been reviewed by the Office of Policy Analysis, U.S. Environmental Protection Agency, and approved for publication. Mention in the text of trade names or commercial products does not constitute endorsement or recommendation for use.

ABSTRACT

The six studies contained in this volume all aim to increase our understanding of the health benefits of air pollution control. However, the link between air pollution and human health remains problematic. One approach to determine such effects is to analyze data on human health affects taken from the real world, uncontrolled, environment and hope that careful statistical analysis will allow one to account for all of the important factors affecting human health so that an unbiased estimate of the effect of air pollution can be determined. This approach is the principal focus of this volume.

The first two studies attempt to determine the relationship between air pollution and mortality. Three of the studies examine morbidity.

In summary, the five statistical studies presented in this volume show: (1) large associations between health and current levels of air pollution are not robust with respect to statistical model specification either for mortality or morbidity and (2) significant relationships, mostly small, do occasionally appear. However it should not be overlooked in light of the rather ambiguous evidence presented in this volume, that all studies to date have only looked for health affects associated with current air pollution exposures, not at any possible association between current health effects and past long term cumulative air pollution exposures.

The final study of this volume attempts to define the type of data which might resolve controversies over the magnitude of air pollution health affects.

CONTENTS

Abstract, ... iii

Figures. vi

Tables. . . . .'. . . . . , . . vii

Chapter I - Introduction. 1

Chapter II - What Have we Learned from Aggregate Data About the Benefits of Air Pollution Control?. 4

Chapter III - Longevity and Air Pollution: A Study Based on Micro Data.. 16

Chapter IV - A Study of Air Pollution-Induced Chronic Illness. 30

Chapter V - Measuring the Benefits From Reduced Acute Morbidity . . . 52

Chapter VI - Air Pollution and Disease: An Evaluation of the NASTwins. 61

Chapter VII - Analytical Priors and the Selection of an "Ideal" Air Pollution Epidemiology Data Set 127

FIGURES

Number		Page
6-1	Major Relationships Examined and Statistically Estimated fortheNASTwins	63
6-2	Conceptual Tradeoff Between Body Capital and Respiration	69
6-3	NAS Twins (Ql) Self-Reported Medical History Questionnaire	77
6-4	NAS Twins (Q2) Residence and Work History	78
6-5	Measuring Benefits from Pollution Reduction Assuming Increasing Costs of Pollution	108
7-1	A Schematic for Air Pollution Health Effects	133

TABLES

Number		Page
2-1	Description of Data and Empirical Estimates	9
3-1	Dependent and Independent Variables Considered intheStudy. . . . . , . . . ,	21
3-2	Description, Mean, and Standard Deviation of the Variables for the "Average Data" Set.	24
3-3	Description, Mean, and Standard Deviation of the Variables for the "Year of Death" Data Set.	25
3-4	The Relationship Between Age at Death and the Relevant Variables -- the "Average Data" Set.	26
3-5	The Relationship Between Age at Death and the Relevant Variables -- the "Year of Death" Data Set. . . .	26
4-1	Complete Variable Definitions	38
4-2	Maximum Likelihood Estimates of Self-Reported Chronic Illness (DSAB).	42
5-1	Health Equations for Men 18-45 Years Old.	57.1
6-1	Age Distribution of National Academy of Sciences Twin Sample - 1967	76
6-2	Definition of Variables	83
6-3	Means and Standard Deviations of Variables.	85
6-4	Correlation Matrix.	86
6-5	Alternative Ordinary Least Squares Regressions with Chest Pain as the Dependent Variable. t-Statistics are in Parentheses	88
6-6	Alternative Ordinary Least Squares Regressions with Severe Chest Pain as the Dependent Variable. t-Statistics are in Parentheses	90

Numb er		Page
6-7	Alternative Ordinary Least Squares with the Incidence of Coronary Heart Attack as the Dependent Variable. t-Statistics are in Parentheses.	91
6-8	Alternative Ordinary Least Squares Regressions with Cough as the Dependent Variable. t-Statistics are in Parentheses	92
6-9	Alternative Ordinary Least Squares Regressions with Shortness of Breath as the Dependent Variable. t-Statistics are in Parentheses.	93
6-10	"t" Statistics on Air Pollution Coefficients, Selected Regression, NAS Twins Data Set . . . . . . , ...	94
6-11	Elasticities of the Incidence Rate of a Symptom with Respect to Air Pollution.	95
6-12	Estimated Annual Per Case Cost of Disease, by Type of Disease, of Cost in 1969 Dollars.	100
6-13	Per Capita Prevalence and Mortality Rates of Specific Diseases in the United States	103
6 - 1 4	The Change in the Total Annual Per Capita Expected Cost of a Symptom Due to a Unit Change in the Pollution Level by Symptom and Disease.	105
6-15	Change in Per Capita Annual Expected Cost of Symptom Given a Change in the Pollution Level	107
6-16	Total Cost Savings, by Symptom, for a 30 Percent Improvement in U.S. Air Quality in 1981 Dollars	109
6-17	Figures used to Calculate the Yearly Consumption of Different Nutrients for the Questionnaire Respondents by Type of Food Consumed and Type of Response Where Appropriate	116
6-18	Figures Used to Calculate Yearly Consumption of Nitrosamines by Questionnaire Respondents by Type of Food Consumed and Questionnaire Response.	119
6-19	Levels of Nutrients and Nitrosamines per Serving byTypeofFood.	120
7-1	Some Factors Which Impact Upon an Assessment of Lung Function in a Population.	140

CHAPTER I

INTRODUCTION

The six studies contained in this volume all aim to increase our understanding of the health benefits of air pollution control. The calculation of health benefits requires both an understanding of how people themselves value health in dollar terms (measured by the willingness to pay concept) and an understanding of air pollution induced health effects. Progress has been made with respect to the former problem. However, the link between air pollution and human health remains problematic. TWO approaches are available for determining the health effects of air pollution. First, animal experiments or, rarely, human experimentation can provide direct evidence in a controlled situation. The second approach is to analyze data on human health effects taken from the real world, uncontrolled, environment and hope that careful statistical analysis will allow one to account for all of the important factors determining human health so that an unbiased estimate of the effect of air pollution can be determined. This latter approach is the principal focus of this volume.

The first three studies attempt to determine the relationship between air pollution and mortality. Chapter 2 examines evidence from data on aggregate mortality rates in sixty U.S. cities and points out the extraordinary difficulty in obtaining a stable, robust statistical relationship between current air pollution levels and current mortality rates. The conventional wisdom holds that a large positive relationship exists between particulate in air and mortality. In Chapter 2, it is demonstrated that this relationship is highly unstable depending on specification of the statistical model used in the analysis. Chapter 3 attempts, using a small sample of data on individual ages at death taken from the Survey on Income Dynamics (1972), to see if, by using disaggregate information on individuals, a more stable and convincing relationship can be obtained. In this small sample of individuals, no significant statistical relationship is obtained between current air pollution levels and longevity.

Three of the studies examine morbidity. Chapter 4 focuses on chronic illness while Chapter 5 focuses on acute illness, where both studies use Survey on Income Dynamics data and data on current air pollution levels. The relationship between chronic illness and air pollution is shown to be potentially large but again very sensitive to model specification. Since little a priori knowledge is available on appropriate model specification, it is impossible to choose between a specification which yields a large impact and one which yields no significant impact. The study of acute health impacts shows, using a particular specification, a small relationship of marginal statistical significance between sulfur oxide and lost work days. Chapter 6 uses an excellent and highly detailed data set on twins collected by the National Academy of Sciences [Hrubec and Neel (1978)1. Of the studies relating to mortality, this one has perhaps the best data and should be capable of detecting even small effects. In fact, a positive but small statistical relationship is shown between air pollution and symptoms of cardiovascular disease such as chest pain. However, the relationship to coronary heart attack is also both quite small and not as strong.

In summary, the five statistical studies presented in this volume show: (1) large associations between health and current levels of air pollution are not robust with respect to statistical model specification either for mortality or morbidity; and (2) statistically significant relationships, mostly small, do occasionally appear.

The final study of this volume, Chapter 7, attempts to define the type of data which might resolve controversies over the magnitude of air pollution health affects. The principal conclusion is that, before a *very* expensive primary data collection effort is undertaken, it would be better to continue statistical modeling of human health effects working with existing data sets, some of which are of fairly high quality. However, all work of this sort should henceforth be built upon explicit physiological and economic models that specify the parameter space. These results can then be used to guide the specification of future primary data collection efforts.

As a final remark which should not be overlooked in light of the rather ambiguous evidence presented in this volume, all studies to date have only looked for health effects associated with current air pollution exposures, not at any possible association between current health effects and long term cumulative air pollution exposures. Thus, it is premature to draw any final conclusions based on existing epidemiological evidence concerning human health and air pollution exposures.

REFERENCES

Survey Research Center. 1972. A Panel Study of Income Dynamics, Ann Arbor, Michigan: Institute for Social Research, The University of Michigan.

Hrubec, Z. and L. V. Neel. 1978. "The National Academy of Sciences National Research Council Twin Registry: Ten Years of Operation," Twin Research: Biology and Epidemiology, New York: Alan R. Liss, Inc.: 153172.

CHAPTER II

WHAT HAVE WE LEARNED FROM AGGREGATE DATA ABOUT THE BENEFITS OF AIR POLLUTION CONTROL?

INTRODUCTION

According to conventional wisdom, the main benefit of environmental regulation is improved health. Thus, research into the benefits of air pollution control has sought primarily to determine the extent to which morbidity and mortality rates decline when air quality improves. Given a knowledge of this relationship, benefits of air pollution regulations can be estimated using the economic analysis of safety programs developed by such investigators as Mishan (1971), Thaler and Rosen (1975), Smith (1974), and Conley (1976). The conceptual framework developed by these authors values small changes in risk using a willingness to pay measure, rather than the lost productivity (or earnings) from early death, and therefore avoids the numerous theoretical problems associated with the latter approach. However, the distinction between these two approaches to benefit estimation reaches far beyond purely theoretical considerations. For similar safety programs, estimates based upon willingness to pay measures are about ten times higher than those based upon productivity changes.[1]

Although progress has been made in valuing the benefits of improved health, the mortality effects of air pollution are less well understood, in spite of the claims of several statistical studies that a clear linkage exists. This paper argues that extraordinary difficulties are present in statistical epidemiology which have yet to be resolved. These difficulties arise in part because of problems in obtaining desirable data. Potential sources of information include first, controlled experimental data from either animal experiments or clinical trials and second, uncontrolled data on human health and exposures in the real world.

Of course, economists have been quick to recognize the similarity of this latter epidemiological problem to many in economics which have been studied using statistical tools such as regression analysis. Use of ordinary least squares to attempt to account for uncontrolled factors and isolate the independent contribution of air pollution to human mortality has become quite popular [see work by Lave and Seskin (1977), McDonald and Schwing (1973), Kneese and Schulze (1977), Crocker, Schulze et al. (1980)1. However, with only a few exceptions, these studies have been unsophisticated in their application of econometric methods and have failed to look for, or cope with, a variety of potentially serious statistical problems.

The plan of the paper is to list a few of these problems in the next section and then to show how these problems can significantly affect estimated effects of air pollution on health using a data set consisting of mortality rates, air pollution levels and other variables for sixty U.S. cities.[2] Comments on policy implications are made in the conclusion.

STATISTICAL PROBLEMS

The aim of this section is to outline some of the major statistical research problems that remain to be overcome in estimating the impact of air pollution on human health. These problems arise largely because the process by which air pollution affects health is not yet completely understood. As a result, any statistical specification of this relationship for the purpose of regression analysis is subject both to uncertainty and question. Most importantly, since the true model is not known with any degree of precision, the power of classical tests of hypotheses regarding the role of air pollution in causing illness or premature death is greatly diminished. To at least some extent, statisticians have faced difficulties of this general nature in virtually all areas of investigation. However, important environmental management decisions regarding air pollution control have been based, in part, upon regression equations where small changes in model specification appear to produce comparatively large changes in implications.

Because theoretical knowledge regarding the connection between air pollution and health is so inadequate, empirical efforts to identify this relationship must be interpreted with caution. Intuitively, there are at least three important types of specification error that should be thoroughly investigated prior to accepting present estimates for policy purposes: (1) errors in functional form, (2) omitted variables, and (3) simultaneity. Clearly, these problems are not an exhaustive list of statistical difficulties in air pollution epidemiology research. Nevertheless, as will be argued momentarily, they do appear to lie at the root of many of the conflicting sets of estimates that have been obtained by other investigations. Each of these three problems will now be considered in turn."

Economic and epidemiological theory provides few insights into the most appropriate functional form for a regression equation used to measure the impact of changes in air quality on human health. This situation is rather unfortunate since the true relationship between health and its determinants may be strongly non-linear. For example, the health consequences of changes in variables such as cigarette smoking, protein consumption, as well as air pollutants are likely to depend not only on the magnitude of the change, but also upon the levels of the variables themselves. Yet little is known about exactly how to specify these functional relationships. The issue of correct function form is important because benefit estimates are frequently obtained from simple equations where a mortality rate (or its natural logarithm) has been regressed on air pollution measures together with other explanatory variables (or their natural logarithms). In particular, these regressions are used to obtain the desired benefit estimates by making hypothetical changes in the air quality variables and then noting the effect on the health measure. Obviously, benefit estimates obtained by this procedure may be seriously biased unless these simple linear or log-linear functional specifications are accurate to a useful degree of approximation.

A second important consequence of the lack of information on the true air quality-health relationship involves the issue of omitted variables. As Theil (1957) has shown, the error of mistakenly excluding variables from an otherwise correctly specified regression equation causes the estimated coefficients on all remaining included regressors to be biased and inconsistent. This issue is not unique to statistical work in the area under study; however, it seems particularly critical here because of apparent conflicts over the empirical determinants of mortality. On the one hand, previous investigations have shown significant adverse health effects resulting from cigarette smoking and certain dietary habits. Nevertheless, when Smith (1975) analyzed thirty-two possible specifications of a regression equation (which are similar to those used by Lave and Seskin (1973)) where the dependent variable was the rate of mortality by SMSA and the explanatory variables were selected from among: (1) median age, (2) percent non-white, (3) population density, (4) temperature, and (5) particulate, little evidence of an omitted variables problem was found to be present. The RESET test, devised by Ramsey (1974), rejected the null hypothesis of a zero mean vector for the disturbance in only five of the thirty-two cases, while the RASET test failed to reject this null hypothesis in all cases. Because these tests were performed at the 10% level of significance and because their results may be unique to the particular data set employed, the appropriate role for other intuitively relevant variables in mortality rate estimating equations legitimately remains the subject of debate. Nevertheless, these results do lend support to the Lave and Seskin estimates of the impact of air pollution on health in the face of charges by other investigators, including Crocker, Schulze et al. (1979), that they have omitted key mortality determinants.

Third, even though the results of Smith's RASET and RESET tests argue to the contrary, the estimation of an appropriately specified air pollution and health relationship may require the use of simultaneous equation estimation methods. Human decision-making may cause the link between these two classes of variables to be considerably more complex than can be captured by a single equation. As an illustration, suppose that increases in medical care are effective in reducing mortality but that mortality rates exert an influence over where medical doctors and others in the health care field choose to locate. In this situation, a medical care variable should be included as an explanatory variable in a regression equation to explain the variation in mortality rates. Simple ordinary least squares estimation, however, may lead to biased and inconsistent estimates of all regression coefficients since the medical care variable would be correlated with the disturbance term even if the number of observations were arbitrarily large. A simultaneous equations estimation technique would be more appropriate in order to explicitly handle the problems created by this correlation.

In addition to the three factors just discussed, two less tractable, but no less important, research problems should be mentioned. First, as discussed by McDonald and Schwing (1973) the variables used to measure air pollutants are often highly correlated with other explanatory variables. Because these pollutants are generated as joint products, in most cases, with other goods produced by the economic system, this situation should not be surprising. If the linear association between explanatory variables is high, separating the independent contribution of each to explaining the variation in mortality rates becomes difficult. McDonald and Schwing proposed a ridge regression estimator as a means of circumventing this problem. Ridge regression methods, however, are not entirely defensible as they represent a rather arbitrary, purely statistical solution to the multicollinearity problem and introduce a bias into the coefficient estimates that would not otherwise be present. (For a more complete critique of ridge regression procedures, see Smith and Campbell, 1980 together with various rejoiners to their paper.) Second, regression models are not highly sensitive and sophisticated research tools, particularly when the data used to estimate them contain measurement error. 'Such models may represent the best statistical tools available to social scientists. Nevertheless, they may not be up to the task of discerning the effect of air pollution on health when, in a correctly specified equation, other explanatory variables may be of much greater importance.

AN EXAMPLE

In this section, two tentative statistical models are presented in order to illustrate the importance of the problems relating to omitted variables and simultaneity that were raised in the previous section. Issues relating to such matters as the choice of functional form and multicollinearity are not explicitly treated here, although they are certainly not less critical subjects for analysis. The first of these models, both of which are estimated using aggregate data on total mortality rates and other variables from sixty U.S. cities, is specified in the equation shown below.

$$MORT = f(NONW, MAGE, DENS, S02X, PART, N02X) \qquad (1)$$

The exact definitions of all variables appearing in this equation, which are similar to those used by Smith and Lave and Seskin, are presented in Table 1. In Equation (l), variations in total mortality rates (MORT) are explained using variables measuring percent non-white (NONW), median age (MAGE), temperature (COLD), as well as the air pollutants (SO2X, PART, and NO2X). Ordinary least squares (OLS) estimates of this equation are presented in the column labeled 1 of Table 1 and t-statistics are presented beneath each coefficient estimate. These findings suggest that SMSAS with more older age residents, more non-whites, and higher air pollution levels (especially in the form of particulate) have, in a statistical sense, significantly higher mortality rates at the 5% level. Examining only this equation, then, leads to the conclusion that air pollution kills people and that appropriate public policy measures should be taken to mitigate this hazard.

Rather different conclusions, however, are obtained from the statistical estimates of the second model. This model is specified in Equations (2) and (3) and the exact definitions of all variables appearing there are given in Table 1.

$$MORT = g(MDPC, NONW, MAGE, DENS, COLD, CIGS, PROT, CARB, SFAT, S02X, PART, N02X) \qquad (2)$$

$$MDPC = h(MORT, INCM, EDUC, S02X, PART, N02X) \qquad (3)$$

Essentially, this structure builds upon Equation (l). Equation (2) explains variations in MORT using variables including NONW, MAGE, and DENS, as well as SO2X, PART, and N02X. But in addition, Equation (2) also allows explicitly for the possibility that mortality rates are affected by cold temperatures

Table 2.1

DESCRIPTION OF DATA AND EMPIRICAL ESTIMATES

Description of Data					Empirical Estimates (t-stat in parenthesis)			
Variable	Year	Units	Mean	SD	NORT (1)	MORT (2)	MDPC (3)	MORT (4)
MORT Total Mortality*	1970	Deaths/1000	11.283	2.16}			5.823 (1.392)	
MDPC Medical Doctors per Capita*	1970	MDs/100,000	162.8	54.2		-.087 (-5.764)		
NONW Nonwhite Population	1969	Fraction	.226	.154	2.997 (2.4o3)	9.996 (6.389)		2.349 (2.365)
MAGE Median Age of Population	1969	Years	28.82	2.74	.573 (8.665)	.789 (13.617)		.626 (11.510)
DENS Crowding in Homes	1969	% > 1.5 persons/room	.022	0.013	12.940 (.881)	49.794 (3.934)		18.217 (1 .447)
COLD Cold Weather	1972	# days temp < 0° C	86.9	47.7		.021 (4.468)		.0175 (3.421)
CIGS Cigarette Consumption	1968	packs/yr/cap	165.8	23.25		.041 (4.693)		.00034 (.526)
PROT Animal Protein Consumption	1965	g/yr/cap	28,128.	1,603.4		.003 (5.032)		.00047 (1.466)
CARB Carbohydrate Consumption	1965	g/yr/cap	123,490.	3,623.0		-.0001 (-2.366)		-.00013 (-).871)
SFAT Saturated Fatty Acids	1965	g/yr/cap	16,315.	976.3		.0016 (4.161)		-.00068 (-2.616)
INCM Median income .	1969	$/yr/household	10,763.	1,060.			.00925 (1.143)	-.000747 (-5.003)
EDUC Education	1969	% > 25 yrs	55.3	7.4			.704 (.616)	-.028 (-.893)
SO2X Sulfer Dioxide	1970	mg/m3	26.92	22.2	.009 (1.059)	-.968 (-4.594)	.070 (.192)	.00118 (.141)
PART Suspended Particulate	1970	mg/m3	102.30	30.11	.011 (2.006)	-.015 (-2.501)	-.514 (-2.085)	.000194 (.0374)
NO2X Nitrogen Dioxide	i969	ppm	.076	.034	1.436 (.271)	-11.081 (-2.332)	-87.228 (-.381)	5.415 (1 .238)
CONSTANT					-7.719	131.48	15.969	7.290
Degrees of Freedom					53	47	53	46
R^2					.692	--	--	.853
Estimation Method					OLS	2SLS	2SLS	OLS

*Predicted values, MORT or MDPC, are employed if these variables are used as explanatory variables in an estimated equation.

(COLD) and by such lifestyle factors as cigarette smoking (CIGS), and diet (PROT, CARB, and SFAT), and by availability of medical care as measured by medical doctors (MDs) per capita (MDPC). Equation (3), then hypothesizes that the location of MDs is determined by total mortality rates, SMSA income (INCM) and education (EDUC) levels as well as by the air quality variables.

Equations (2) and. (3) are simultaneous in that variations in MORT are determined, in part, by variations in MDPC and vice-versa. Due to this fact, and because under the order condition both equations appear to be identified, two stage least squares (2SLS) is used as an estimation method. The estimates of these two structural equations are given in columns labeled 2 and 3 of Table 1. With the exception of the coefficients on the air pollution variables, estimates of the slope parameters in Equation (1) possess signs that might be expected on intuitive grounds. Increases in MDPC and in CARB contribute significantly to reductions in mortality rates, while colder SMSAS with more older age residents, more non-whites, more crowded housing conditions and where more cigarettes are consumed tend to have higher mortality rates. These results suggest that holding constant the linear influence of medical doctors per capita, lifestyle variables measuring such factors as smoking and dietary habits exert a significant influence on total mortality rates; a finding that is of interest since variables of this type were ignored in specifying Equation (l). On the other hand, the statistically significant but negative coefficients on the air pollution variables are rather more of a puzzle and cannot be completely explained. Nevertheless, a partial account of why this anomalous result has occurred will be offered momentarily. In the meantime, consider the estimates of the slope parameters of Equation (3). According to these estimates, all but one of which are not statistically significant at conventional levels, medical doctors apparently avoid locating in SMSAS where particulate levels are high.

Additional insights into these results can be obtained by examining the estimates of the reduced form equation for MORT, which are shown in the column labeled 4 of Table 1. As indicated in the table, these estimates were obtained by applying ordinary least squares to an equation where MORT was specified to be a function of all exogenous variables in the structural model presented previously. There are two aspects of these estimates that are particularly worth noting. First, the estimates of the reduced form coefficients, unlike the structural coefficients, do not hold constant the linear influence of medical care and are interpreted as total, rather than partial, derivatives. In other words, the structural coefficients do not fully capture the fact that medical care may ameliorate the negative health effects of cigarette smoking, cold weather, crowded living conditions, and so forth. This ameliorative effect can only be determined by comparing the reduced form to the structural form coefficients. As is evident, such a comparison reveals that the coefficients on the socioeconomic and lifestyle variables are all smaller in the reduced form than in the structural form; a result suggesting that some ameliorative effects of medical care may indeed be present. Second, in the reduced form mortality equation, the coefficients on the air pollution variables are positive. How can this result be explained? Although increased medical care would appear to reduce total mortality rates, doctors, according to the structural equation estimates, prefer not to live in polluted areas. Consequently, the reduced form coefficients, which take this behavior into account, are larger than their counterparts in the structural form. This observation, clearly, does not explain why the structural air pollution coefficients are negative. However, it does suggest that reduced form expressions will allow the net effects to be estimated.

CONCLUSION

Existing statistical work on the mortality effects of air pollution has been interpreted to imply that control of stationary sources such as power plants (which emit SO_2 and particulate) is justified while auto emission controls (particularly those for nitrogen oxides) are unjustified. These conclusions may be unwarranted for two reasons. First, as shown in the preceding section, the estimated effects of air pollution on human health are highly sensitive to model specification. With little or no a priori theoretical rationale for choosing one specification over another, a determination of the true health effects of air pollution is impossible. Future research, with primary data that is both collected specifically for the purpose of analyzing the health effects of air pollution and aimed at coping with the kinds of statistical problems identified here, may provide more convincing estimates. At the present time, however, relatively little is known about the effects of long term low-level air pollution exposures on human mortality; certainly not enough to make benefit projections for policy purposes.

Second, the really important benefits from air pollution control may actually lie in the non-health area. For example, a recent study of the Los Angeles basin suggested that a 30% reduction in ambient pollution levels (principally nitrogen oxides and related oxidant) would be worth nearly one billion dollars per year to local residents (Brookshire et al. 1980). This study, using both a traditional hedonic property value study and survey questionnaires, concluded that a major fraction of perceived benefits was derived from the aesthetic (visibility, and quality of life) benefits of reduced air pollution. Similarly, studies of the benefits of air pollution control in recreation areas such as the national parklands of the southwest suggest that visibility and related non-health benefits are of principle concern. While supposed effects of air pollution on human mortality provide decisionmakers with an easy justification for control policies (often on ethical rather than economic grounds), economists ought to be concerned with all sources of benefits from pollution control on efficiency grounds. Serious doubt over the health effects of air pollution implies that less emphasis should be placed on health effects in making policy decisions.

REFERENCES

1 For example, Lave and Seskin (1977) use about $30,000 as an average value of a life saved in increased productivity based on the work of Rice (1968). In contrast, Crocker, Schulze et al. (1979) use $340,000 as the willingness to pay for an expected life saved based on the work of Thaler and Rosen (1975).

2 For a more complete examination of this data set see Schulze, Ben-David, Kneese and Pazand, "Mortality, Medicine, and Lifestyle," mimeo, University of Wyoming, January 1980.

BIBLIOGRAPHY

Brookshire, D., R. d'Arge, W. Schulze and M. Thayer, "Experiments in Valuing Public Goods," in V. Kerry Smith, cd., Advances in Applied Microeconomics, forthcoming.

Conley, B., "The Value of Human Life in the Demand for Human Safety," American Economic Review, 66(March 1976), p. 5457.

Crocker, T., W. Schulze, S. Ben-David and A. V. Kneese, Methods Development for Assessing Air Pollution Control Benefits, Volume I EPA-600/5-79-00/a, February 1979.

Kneese, A. V. and W. Schulze, "Environment, Health and Economics - The Case of Cancer," American Economic Review, 67(February 1977), p. 26-32.

Lave, L. B. and E. P. Seskin, "An Analysis of the Association Between U.S. Mortality and Air Pollution," Journal of the American Statistical Association, 68(June 1973), p. 284-90.

Lave, L. and E. Seskin, Air Pollution and Human Health, Baltimore, 1977.

McDonald, G. C. and R. C. Schwing, "Instabilities of Regression Estimates Relating Air Pollution to Mortality," Technometrics, 15(August 1973), p. 463-81.

Mishan, E. J., "Evaluation of Life and Limb: A Theoretical Approach," Journal of Political Economy, 79(July/August 1971), p. 687-705.

Ramsey, J. B., "Classical Model Selection Through Specification Error Tests," in Paul Zarembka, cd., Frontiers in Econometrics, 1974.

Schulze, W., S. Ben-David, A. Kneese and R. Pazand, "Mortality, Medicine, and Lifestyle," mimeo, University of Wyoming, January 1980.

Smith, G. and G. Campbell, "A Critique of Some Ridge Regression Methods," Journal of the American Statistical Association, 75(March 1980), p. 74-81.

Smith, R. S., "The Feasibility of an 'Injury Tax Approach' to Occupational Safety," _Law and Contemporary Problems,_ 38(Summer-Autumn 1974), p. 730-744.

V. K. Smith, "Mortality - Air Pollution Relationships: A Comment," _Journal of the American Statistical Association,_ 70(June 1975), p. 341-43.

R. Thaler and S. Rosen, "The Value of Saving a Life: Evidence from the Labor Market," in Nestor E. Terleckyj, cd., _Household Production and Consumption_, New York, 1975.

H. Theil, "Specification Errors and the Estimation of Economic Relationships," _Review of the International Statistical Institute,_ 25(1957), p. 41-51.

CHAPTER III
LONGEVITY AND AIR POLLUTION
A STUDY BASED ON MICRO DATA

INTRODUCTION

The health effects of air pollution has been intensively studied and discussed by various scientists and researchers in the recent years. Many of such studies have found statistically significant positive relationships between air pollution and morbidity [Fishelson and Grove (1978); Sterling, et al. (1967) Sterling, et al. (1969)] as well as mortality [Kneese and Schulze (1977); Koshal and Koshal (1974); Lave and Seskin (1977); Schwing and McDonald (1976)]. Lave and Seskin are among the scientists who have conducted an extensive research on the subject matter. The result of their three consecutive studies, utilizing 1960, 61, and 69 aggregate data for several U.S. cities, as well as an intensive review of the related studies appear in the publication entitled Air Pollution and Human Health [Lave and Seskin (1977)]. This publication strongly suggests that there exists a significant positive relationship between air pollution and mortality. Schulze, et al. have conducted similar studies on the human health effects of air pollution [Kneese and Schulze (1977)]. According to their most recent study the health effects of air pollution is indirect (U.S. Environmental Protection Agency EPA600/579001a) . In this study Schulze, et al. suggest that air pollution is one of the factors affecting the location decision by physicians, in the sense that doctors consider air pollution a disamenity, hence avoid polluted areas if possible. Furthermore, they reason that the supply of physicians undoubtedly affects the mortality rate by decreasing the probability of a premature death event occurring in cases of emergency, and/or increasing longevity through providing health services. The study reasons that if air pollution discourages physicians from locating in a specific area, and if scarcity of doctors increases the mortality rate, then excluding the supply of doctors as an explanatory variable from the epidemiological model leads to observing a strong positive relationship between air pollution and mortality. Schulze, et al. conclude that although air pollution adversely affects human health, the strong positive relationship between air pollution and mortality, as observed in the statistical studies, is misleading. In other words the health effect of unavailability of health services rendered by physicians may dominate the adverse health effects of air pollution.

The Institute for Social Research of the University of Michigan has conducted a survey entitled A Panel Study of Income Dynamics (from now on referred to as the Michigan Study) in which about 5000 families, chosen at random from 50 states of the United States, have been interviewed from 1968-1976 [Institute for Social Research (1977)]. The Michigan Study has interviewed the families in the sample on an annual basis and has collected numerous information such as age, sex, race, state and county of residence at the time of the interview as well as childhood, parent's economic status and education, current and previous employment, distance to work, driving habits, income, education, life style, eating habits, health insurance, illness, physical condition, and several other relevant facts for the head of each family. During the survey period, the head of some of the families in the sample has died (or separated, or otherwise moved away) , and hence a sub-sample in the Michigan Study is created. This sub-sample (and hence the Michigan Study) provides an excellent chance of examining the possible relationship between air pollution and longevity. The Michigan Study provides detailed information about the length of life, background variables (such as the race and sex of the sample member, and the parent's economic condition when the sample member was growing up) , current variables (such as the size of the city of residence at the time of interview, distance to work, education, per capita cigarette and alcohol consumption), and health variables (such as insurance coverage, illness history, annual income and quality of air). A great deal of this information is difficult to acquire under normal circumstances. The present study utilizes the aforementioned sub-sample of the Michigan Study to investigate the possible relationship between air pollution and longevity.

DESCRIPTION OF DATA

The Michigan Study provides a wide range of information about the head of the families in the sample who have died during the survey period, 19681976. The sub-sample, the set of interviewees who have died during the aforementioned time period, consists of 568 observations. From now on the aforementioned sub-sample of the Michigan Study is the focus of attention. The Michigan Study has not attempted to explore the cause of death for the sample members. Using the information compiled in the Michigan Study, it is possible to establish a statistical relationship between the age at death and several relevant variables that may fall into three broad categories:

1) background variables: such as the race and sex of the sample member, and the parent's economic condition when the sample member was growing up;

2) current variables: such as the size of the city of residence at

the time of interview, distance to work education, per capita cigarette and alcohol consumption;

3) and health variables: such as insurance coverage, illness history, annual income, and quality of air.

In order to implement such a study, it was felt necessary to look closely into the data set. It was soon realized that the data set as it stood was not suitable for a meaningful statistical analysis. It was observed that age, race, sex, city size as well as state and county of residence when the sample member was growing up, to mention only a few variables, changes several times for most of the sample members. Following a careful investigation of the data set the reason for such a disturbing occurrence was discovered. The following example should shed light into the source of this problem. Suppose Mr. X has been the head of a family and has been interviewed from 1968 through 1970 as one of the sample members of the Michigan Study. Suppose Mr. X dies in 1970 and Mrs. Y replaces him. From 1971 no more information is collected for Mr. X and all variables pertaining to Mr. X takes on a zero value for the remainder of the survey period. Mrs. Y has not been interviewed as the head of this particular family for the years 1968 through 1970, hence no information about Mrs. Y is available for this time period. Information collected for Mr. X for the years 1968 through 1970 is assigned to Mrs. Y. From 1971 onward, information about Mrs. Y is properly collected. So far two observations have been created from only one head of the family, Mr. X. One observation contains information about Mr. X alone from 1968 through 1970, another observation contains information about Mr. X from 1968 through 1970 and information about Mrs. Y from 1971 onwards. Now suppose Mrs. Y dies in 1972 and Mr. Z takes on her responsibility as the head of the family from 1973. According to the procedures adopted by the Michigan Study, no more information about Mrs. Y is compiled and the variables pertaining to Mrs. Y takes on a zero value for the remainder of the survey period. In the meantime a new observation is created, namely Mr. Z, which contains information about Mr. X for the years 1968 through 1970, information about Mrs. Y for the years 1971 through 1972, and information for Mr. Z for the years 1973 to the year he died. If Mr. Z dies before 1976 and is replaced by, say, Miss W, then yet another observation is created which would contain information about Mr. X, Mrs. Y, Mr. Z and Miss W. Theoretically speaking, one observation could have information about nine different individuals. If the individuals in one observation are numbered from 1 to 9, then information about individual #1 could appear nine times in the data set, eight times for individual #2, seven times for individual #3, . . ., and once for the individual #9. Working with such a data set could provide misleading results. Obviously, before any reliable statistical study could be conducted, the data set had to be cleaned up and a procedure need be adopted to compile a new data set such that the information for each individual appears only once in the data set. One of the possible solutions to the existing problem is to determine the year in which the sample member has died and then choose the value of the relevant variables at the year of death. Accordingly, a data set may be created which would have 568 independent observations with no repetition. There exist two major difficulties with this procedure:

1) Not all the variables that reveal important information have been asked during the entire nine years of the survey period. For instance, the question "whether or not the interviewee has been disabled" has been asked only in 1968 and 1976. The question "whether or not he has had a disabling illness in the past" has been asked only in 1968. The question about the trend of disability has been asked in the years 1970 through 1975. The question "whether the individual has been covered by any health insurance" has been asked in the years 1969 through 1972. The question about the amount of money spent on cigarettes and alcohol has been asked only the years 1970 through 1972. These are but a few examples. Therefore, if this procedure is adopted, information about very important variables in the year of death may not be available, simply because the question has not been asked in that year and hence several observations may have to be deleted.

2)' More importantly, since the survey is about the individuals, the value of a variable for a given year may be exceptionally low or high. For instance, income of an individual at the year of death, or the value of any other relevant variable may be lower or higher than usual for a variety of reasons. Therefore accepting this unusual level of income as an independent variable and exploring its affect on the dependent variable could bring about biased result. Hence it may be desirable to know the value of the relevant variable for more than one year and use their average in the statistical model so that the study would be statistically unbiased and hence reliable.

For the aforementioned reasons it was decided to only choose the observations that provide information for a specific individual for at least two consecutive years in the survey period. The age variable was used as the prime determinant. It is obvious that if the age variable for one observation does not consistently increase by one unit during the survey period, that observation contains information about more than one individual. To make certain that each observation contains information about a specific individual, age, sex, race, and the city size when growing up were utilized as control variables. Following this procedure, the sample size was reduced from

568 to 153. The 153 observations in the new smaller data set are virtually independent of one another in the sense that each observation contains information about one specific individual, furthermore, each individual has been interviewed at least two consecutive years during the survey period and hence for each variable of interest there may exist information for at least two years (given the relevant question had been asked in the years the individual has been interviewed) such that their average could be employed in the statistical study. The data set thusly compiled will be referred to as the "average data" set.

Table 1 lists the dependent and the independent variables that were chosen from the information available in the Michigan Study based on the thought that they might have significant relationships with the dependent variable: age at death. Meanwhile, the methodology for narrowing down the several-year-information for each variable into one unique number is explained.

The constructed "average data" set, as previously described, consists of 153 observations which are independent of one another in the sense that each observation contains information about one specific individual. But since not all questions had been asked in all nine years of the survey period, several observations in the average data set do not provide information about some of the relevant variables considered in this study. Hence, at the final stage, before adding the air quality variables, the average data set was reduced to 114 observations. The last stage of the study was to incorporate the air quality variables into the statistical model. For privacy purposes, only the county of residence of the sample members is provided by the Michigan Study. The mean annual concentration of suspended particulate and sulfur dioxide for counties during the years 1968 to 1976 was obtained and added to the "average data" set [U.S. Environmental Protection Agency (1968)-(1976)]. Unfortunately, air quality information in the survey period was available only for some of the counties. Therefore, after the air quality variables were added to the average data set, more observations had to be deleted and the new average data set was further reduced to 51 observations. Based on this data set a statistical model is developed and a relationship between the age at death and several relevant variables is established. The results of the statistical model are discussed at the end of the next section; but since the size of the "average data" set at the final stage turned out to be rather small, it was decided to compile another data set hoping it would contain a larger number of observations. It was decided to choose the value of the relevant variables at the year of death from the original sub-sample with 568 observations. This

TABLE 3.1

DEPENDENT AND INDEPENDENT VARIABLES CONSIDERED IN THE STUDY

I - Dependent variable:

Age at death - Age of the individual at the time of death.

II - Independent Variables:

A - Background variables:

1 - Sex: 0 = male, 1 = female

2 - Race: 0 = white, 1 = non-white (includes Puerto Rican, Mexican, Cuban, and others).

3 - Region when growing up:* 1 = Northeast, 2 = North Central 3 = South, 4 = West, 5 = Hawaii, Alaska, 6 = all foreign countries, 9 unknown.

4 - City size when growing up:* 1 = farm, 2 = small town, 3 = large city, 4 = other, different place.

5 - Parent's economic condition when growing up: 0 = poor, 1 = well off. Mode of observations was chosen.

Variables number 1-4 in group A were used as control variables, hence no discrepancy existed.

B .- Current and, health variables:

6 - Distance to a city of 50,000 or more at the time of interview:* 1 = under 5 miles, 2 = 5-14.9 miles, 3 = 15-29.9 miles, 4 = 30-49.9 miles, 5 = 50 miles or more. Mode of observations was chosen.

7 - Miles to work:* 00 = none, neither drives nor has car pool, unemployed, retired, student, etc. 01 = one mile or less, 02 = two miles, . . ., 98 = 98 miles or more, 99 = N/A. Average of observations (excluding 99) was chosen.

8 - Miles driven per year:* 00 = N/A, none, no car, XXXXX = actual miles driven, 99998 = 99998 miles or more, 99999 = unknown. Average of observations (excluding 99999) was chosen.

9 - Whether disabled: 1 = yes, complete limitation on work, 2 = yes, severe limitations on work, 3 = yes, some limitation on work, 4 = yes, no limitation on work, 5 = no, 7-9 = N/A. 1-3 was assigned 1; 4, 5 were assigned 0; 7, 9 = no information available. Mode of observation (excluding 7, 9) was chosen.

Table 3.1 (continued)

10 - Trend''of' disability: 1 = better, 3 = stays the same, 4 = fluctuates, 5 = worse, 9 = N/A, unknown, 0 = inap. (no disability), 1, 3 were assigned 0; 4, 5 were assigned 1. Mode of observations (excluding 9) was chosen. Weight given to more recent observations.
11 - Number of hours ill per year:* 0000 = none, XXXX = actual hours of illness, 9999 = 9999 hours or more. Average of observations was chosen.
12 - Whether covered by health insurance: 1 = yes, 0 = no. Mode of observations was chosen.
13 - Education:* 0 = cannot write or read, 1 = 0-5 grade, 2 = 6-8 grade, 3 = 9-11 grade, 4 = 12 (high school), 5 = 12 grade plus nonacademic training, 6 = college but no degree, 7 = college B.A. , no advance degree, 8 = college and advanced or professional degree, 9 = N/A, unknown. Mode of observations (excluding 9) was chosen.
14 - Total family money income:* Average of observations was chosen.
15 - Number of adults in the family:* Average of observations was chosen.
16 - Number of children in the family:* Average of observations was chosen.
17 - Per capita average income: 14/(15+16)
18 - Amount of money spent on alcohol per family.* Average of observations was chosen.
19 - Amount of money spent on cigarettes per family.* Average of observations was chosen.
20 - Per capita alcohol consumption: 18/15.
21 - Per capita cigarette consumption: 19/15.

*Variable classifications as stated in Michigan Study.

new sample, referred to as the "year of death data," consists of 170 observations and contains information about the relevant variables that are included in the model. This data set is also consistent and the observations are independent of one another since the value of the variables at the year of death has been chosen. The air quality information was collected from the same source as in the case of the "average data" set. To include as many observations as possible in the statistical model, the value of the air quality variables at the year of death was chosen. If air quality variables were not available at the year of death, the value of the air quality variables were not available at the year of death, the value of air quality variables for the year(s) prior to death was chosen. In cases where air quality information for the year of death and year(s) prior to death was not available, the value of air quality variables for the year(s) after death was chosen. Similar procedure has been employed for the "average data" set. In both samples the air quality information for about 75% of the observations are for the year of death (years the sample member has been interviewed for the average data set) about 15% for the year(s) prior to death, and about 10% for the year(s) after death. Following this procedure when air quality variables are included the year of death data set sample size reduces to 63 observations, which contains 12 observations more than the "average data" set.

THE STATISTICAL MODEL

The statistical model used in this study tests the hypothesis that longevity is closely related to background, current, and health variables as discussed in the introduction section. It is hypothesized that age at death is affected by background factors such as sex, race, geographical region and the parent's economic condition when the sample member has been growing up; the current factors such as the size of the city of residence, distance to work, education, cigarette and alcohol consumption; and the health factors such as health insurance coverage, illness history, income, and the quality of air. A series of regression equations have been obtained (for both data sets previously discussed). Careful investigation of the individual regression equation has been the basis for the decision on the final form of the regression equations. Table 2 reports the description of the variables considered in this study and their mean and standard deviation for the "average data" set. Table 3 provides similar information for the "year of death" data set. The regression equations, in their final form, are reported in Tables 4 and 5. Table 4 reports the result of the study when the "average data" set is utilized, Table 5 reports the result of the study when the "year of death" data set is utilized. Each table contains two equations. Equation one is the statistical model in its final form when air quality variables are included. Since the size of both data sets at the final stage turned out to be rather small, it was decided to increase the sample size by not checking for air

TABLE 3.2

DESCRIPTION, MEAN, AND STANDARD DEVIATION
OF THE VARIABLES FOR THE "AVERAGE DATA" SET

Description of the Variables	Mean	Standard Deviation
Age at death (years)	52.92	14.08
Race	.65	1.04
Distance to a city of 50,000 people or more	1.98	.84
Annual hours ill	120.22	185.17
Miles to work	3.86	5.5
Health insurance	.82	.39
Education	3.39	1.92
Education squared	15.12	14.45
Per capita expenditure on alcoholic beverages ($)	44.43	73.05
Per capita expenditure on cigarettes ($)	42.12	59.28
Mean annual concentration of suspended particulate in the air (PPM)	90.47	24.34
Mean annual concentration of sulfur dioxide in the air (PPM)	45.39	48.78

TABLE 3.3

DESCRIPTION, MEAN, AND STANDARD DEVIATION OF
THE VARIABLES FOR THE "YEAR OF DEATH" DATA SET

Description of the Variables	Mean	Standard Deviation
Age at death (years)	51.25	14.78
Distance to a city of 50,000 population or more	1.95	.87
Annual hours ill	149.40	336.74
Miles to work	3.52	6.35
Health insurance	,67	.47
Education	3.52	1.73
Education squared	15.36	13.37
Per capita annual expenditures on alcoholic beverages ($)	53.71	113.21
Per capita annual expenditures on cigarettes ($)	47.32	78.34
Mean annual concentration of total suspended particulate in the air (PPM)	98.25	26.79
Mean annual concentration of sulfur dioxide in the air (PPM)	16.16	11.91

TABLE 3.4

THE RELATIONSHIP BETWEEN AGE AT DEATH AND THE RELEVANT VARIABLES - THE "AVERAGE DATA" SET

(t-statistics in parenthesis)

Age at death	Race	Distance to a major city	Miles to work	Education	Education squared	Per capita annual alcohol consump-tion	Per capita annual cigarette consump-tion	Annual hours ill	Health insurance	Mean annual concen-tration of suspended particulate in the air	Mean annual concen-tration of sulfur dioxide in the air	Constant	R^2	Sample size
Age	1.9 (.93)	-1.2 (-.5)	-.2 (-.5)	-.4 (-.1)	-.08 (-.15)	-.04 (-1,3)	-,02 (-.37)	-.03 (-2.2)	2.7 (.5)	.04 (.3)	-.001 (-.15)	57. ? (4.3)	.26	51
Age	-1.2 (-.9)	.8 (.9)	-.4 (-2.1)	-2.6 (-1.1)	,28 (.9)	-.03 (-1.2)	-.03 (-1.3)	-.03 (-3.0)	-.5 (-.17)	′-		63.8 (11.5)	″23	′ 14

TABLE 3.5

THE RELATIONSHIP BETWEEN AGE AT DEATH AND THE RELEVANT VARIABLES - THE "YEAR OF DEATH" DATA SET

(t-statistics in parenthesis)

Age at death	Race	Distance to a major city	Miles to work	Education	Education squared	Per capita annual alcohol consump-tion	Per capita annual cigarette consump-tion	Annual hours ill	Health insurance	Mean annual concen-tration of suspended particulate in the air	Mean annual concen-tration of sulfur dioxide in the air	Constant	R^2	Sample size
Age	-10.3 (-2.5)	.35 (.15)	.11 (.35)	-.97 (-.22)	-.17 (-.3)	-,007 (-.42)	-.07 (-2.53)	.006 (1.0)	-4,8 (-.92)	-.006 (-.07)	-.05 (-.29)	68.3 (5.4)	.28	63
Age	-6.8 (-4.4)	,82 (1,12)	-.18 (-1.3)	-7.9 (-3.5)	.88 (3.0)	-,01 (-1.0)	-.03 (-2.1)	.001 (!47)	-1.7 (-.69)	--	--	71.0 (14.5)	.24	162

quality variables and observe the sensitivity of the model. Therefore, equation two is identical with equation one but it contains larger numbers of observations by not checking for air quality variables, and hence excluding the air quality variables from the regression equation. The Ordinary Least Square technique has been utilized in obtaining all regression equations. Careful analysis of the regression equations in Tables 4 and 5 leads to the following deductions. Race, among background variables, has a significant inverse relationship with life span of the sample members in this study. This result is in agreement with the existing statistics that whites have a longer average life span than non-whites.

Among current variables, distance to a major city is positively related to the age at death except in equation one of Table 4. The relationship is not generally significant except for equation two of Table 5 where this variable is almost significantly related to life span. This result is also in agreement with existing statistics that rural populations live longer, on the average, than the urban populations. Miles to work is inversely related to age at death, suggesting that people who commute to work have a shorter life span as the risk of having an accident increases with an increase in the commuting distance. According to equation two of Table 4, this variable is significantly related to the age at death. Education has an inverse relationship with longevity. The relationship is strongly significant according to equation two of Table 5 in which education squared has a significant positive relationship with life span. The indication is that as education increases to about grade 12 (high school) life span decreases, but with higher education (past high school) life span increases. According to equation two of Table 5: longevity = 6.8 age + .82 distance to major city .18 miles to work 7.9 education + .88 (education)^2 - .01 alcohol consumption .03 cigarette consumption + .001 annual hours ill 1.7 health insurance + 71. The minimum life span is associated with education = 4.46. According to Table 1, this figure refers to a level of education between a high school graduate and a high school graduate with nonacademic training. Therefore it may be concluded that college education increases longevity whereas elementary and high school education has an inverse effect on life span. This finding may be justified by observing the characteristics of the existing job markets. College education increases the chance of acquiring well-paying, less risky jobs. Furthermore, more risky jobs require a certain type of skill which may require education beyond elementary level. Therefore, observing a binomial relationship between education and longevity with minimum life span associated with high school graduate level may not be far from reality. Annual per capita consumption of alcohol and cigarettes are inversely related to longevity. Furthermore, consumption of cigarettes is significantly related with age at death, as indicated by equations one and two of Table 5; which, quite expectedly indicates that cigarette consumption decreases life span.

Among health variables illness is inversely related with longevity and the relationship is significant (Table 4). According to Table 5 the relationship is positive, but insignificant. This finding indicates that illness measured as the average number of hours ill over several consecutive years is the proper measure of illness rather than the number of hours ill at the year of death. Health insurance coverage is inversely related to longevity--an unexpected result (except for equation one of Table 4), but the relationship is totally insignificant. Air pollution (as measured by total suspended particulate and sulfur dioxide) is inversely related with longevity (except for suspended particulate in equation one of Table 4); however, the relationship is not significant.

CONCLUSIONS

The present study investigated the effect of several relevant variables on longevity. Based on a sub-sample of the Michigan Study (a Panel Study of Income Dynamics conducted by the Institute for Social Research of the University of Michigan) two data sets were constructed consisting of the age of the individuals at the year of death and several explanatory variables expected to be related with longevity based on the existing epidemiological studies. Careful investigation of the several Ordinary Least Square regression equations which included different combinations of the explanatory variables lead to the final form of the regression equations reported in Tables 4 and 5. Based on the results of this study, it can be concluded that air pollution, although inversely related to age at death, does not significantly affect longevity. It can also be concluded that education and consumption of alcohol and cigarettes have a stable relationship with longevity since the direction of relationship is consistent in the two equations of the two data sets. Distance to a major city, miles to work, and health insurance are not stable variables affecting longevity. It can also be concluded that longevity increases as education goes beyond high school and also as education stops short of graduating from high school. It may also be concluded that illness measured as the average illness for several consecutive years is a far better health measure than illness at the year of death. Similar reasoning applies to the consumption of alcoholic beverages; however, race and cigarette consumption are more significantly related to longevity if their value at the year of death is included in the study. Finally, considering the inverse relationship between health insurance and longevity, it can be concluded that illness is an endogenous variable since illness decreases an individual’s chance to purchase health insurance and the lack of health insurance shortens an individual’s life span. Therefore, the statistical model of this study may be improved by developing a two-stage model in which illness is an endogenous variable affected by such variables as income, education, race, and age.

BIBLIOGRAPHY

Fishelson, G. and P. Grove, 1978. "Air Pollution and Morbidity: SO Damages, " Journal of Air Pollution Control Association, Vol. 28, No. 8: 785-789.

Institute for Social Research. 1977. "A Panel Study of Income Dynamics," The University of Michigan, Ann Arbor, Michigan.

Kneese, A. V. and W. D. Schulze, 1977. "Environment, Health, and Economics: The Case of Cancer," American Economic Review, Vol. 67, No. 1: 326-332.

Koshal, R. K. and M. Koshal, 1974. "Air Pollution and the Respiratory Disease Mortality in the U.S.: A Quantitative Study," Social Indicator Research, 1(3): 263.

Lave, L. B. and E. P. Seskin, 1977. Air Pollution and Human Health, Baltimore: Johns Hopkins University Press.

Schwing, R. C. and G. C. McDonald, 1976. "Measures of Association of Some Air Pollutants, Natural Ionizing Radiation and Cigarette Smoking With Mortality Rates," Science of Total Environment 5: 139-169.

Sterling, T. D., S. V. Pollack, and J. J. Phair, 1967. "Urban Hospital Morbidity and Air Pollution, A Second Report," Archive of Environmental Health, Vol. 15: 362-374.

Sterling, T. D., S. V. Pollack, and J. Weinkam, 1969. "Measuring the Effect of Air Pollution on Urban Morbidity," Archive of Environmental Health, Vol. 18: 485-494.

United States Environmental Protection Agency, 1968-1976. Air Quality Data, Annual Statistics.

United States Environmental Protection Agency, Methods Development for Assessing Air Pollution Control Benefits, V. I. Experiments in the Economics of Air Pollution Epidemiology , EPA-600/5-79-001a.