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.

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

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.

$$MORT = g(MDPC, NONW, MAGE, DENS, COLD, CIGS, PROT, CARB, SFAT, S02X, PART, NO2X) \quad (2)$$

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

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 (2.403)	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)
N02X Nitrogen Dioxide	1969	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 Micro-economics, 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. 11 - 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 consumption	Per capita annual cigarette consumption	Annual hours ill	Health insurance	Mean annual concentration of suspended particulate in the air	Mean annual concentration 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)		

TABLE 3.5

THE RELATONSHIP 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 consumption	Per capita annual cigarette consumption	Annual hours ill	Health insurance	Mean annual concentration of suspended particulate in the air	Mean annual concentration 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 (1.47)	-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.

CHAPTER IV
A STUDY OF AIR POLLUTION-INDUCED CHRONIC ILLNESS

INTRODUCTION

At the time of the national awakening about environmental issues that occurred in the late 1960's, a great deal of public and scientific attention was focused on statistical relationships between air pollution and human health. While this research was undertaken with a large measure of academic curiosity, a major impetus was provided by Federal government agencies, such as the United States Environmental Protection Agency and its predecessors. The motivating factor for this agency encouragement was a laudable desire to establish scientific evidence for regulations designed to mitigate any detrimental health consequences of air pollution. For a time in the mid-1970's, the subject, though continuing to be discussed in scientific councils, did not capture much public attention, perhaps because of substantial reductions in the ambient concentrations of several common air pollutants. However, with the immediate threat that switching from oil and natural gas to coal fuels poses to the progress of a decade in controlling air pollution, the aforementioned statistical relationships are again a subject of public as well as scientific scrutiny.

In this paper, we assess the extent to which exiting epidemiological research can be interpreted as statistically demonstrating a relationship between air pollution and human health status. We also present some additional statistical research of our own. The next section is a critical review of the methodological underpinnings of existing research in air pollution epidemiology. So as not to exempt our previous work from this critical review, we devote a third section to self-appraisal. A fourth section presents some new empirical results meant to respond to several of the faults we confessed in the third section. The two concluding sections summarize what we think we have thus far learned and make some suggestions for future research.

A CRITICAL REVIEW OF OTHERS' WORK

Much of the recent work in air pollution epidemiology has focused upon

estimation of some version of the following expression:

$$H_i = a + bP_i + cX_i + u_i, \quad (1)$$

where H is a measure of morbidity or mortality, P is a measure of pollution, X is a set of other variables thought to influence health status, u is an error term that captures the effects of unmeasured influences upon health status, i indexes the individuals or groups of individuals in a sample, and a, b, and c are parameters to be estimated. Epidemiological work of this sort, a large part of which has been done be economists, presumes that there exists a distribution across individuals of tolerances to air pollutants and that there exist some individuals for whom any air pollution exposures whatsoever will trigger a decline in health status. This perspective may be contrasted with another, common to many epidemiological studies originating in the biomedical disciplines and sanctified in existing Federal clean air legislation, which posits a positive level of air pollution below which no individual will suffer a decline in health status[1].

Two recent empirical applications of the latter perspective are Morris, et al. (1976) and Bauhuys, et al. (1978). Inspired by the principles of experimental design, the researchers in each of these studies selected two communities similar in most respects other than air pollution. Using analysis of variance techniques, statistically significant differences in health status between the populations of the communities were then sought. Whether or not these differences were found, toxicological evidence from laboratory studies was then cited to provide a basis for rejecting or failing to reject air pollution as a cause of the difference. Many of the cited laboratory studies are, in principal, structured in the same fashion as the epidemiological studies; that is, the experimenter takes a treatment group and a control group of similar individual organisms and increases the pollution exposures of the treatment group until a decline in health status is observed. The pollution level at which this decline is first observed is then said to be the threshold at which pollution is universally unhealthy.[2]. Practitioners of this perspective generally agree that most substances commonly termed air pollutants can have deleterious human health effects. The controversies among them erupt over the threshold pollution levels at which these effects emerge and whether these threshold levels are found in everyday human environments. Because the methods provide no information on the magnitudes of any effects that do exist, the controversies are limited to questions on the statistical determination of the existence of an effect.

Unless all factors that contribute to differences in health status across individuals and locations can be controlled, the weaknesses inherent in empirical applications of the above perspective are apparent. In particular,

statistically significant differences between the health states of two groups of individuals may not be observable because the contributions of air pollution to the true differences are overwhelmed by uncontrolled factors. Any perceived threshold is then more a matter of experimental design rather than of effect: perception of 'where the threshold lies will differ with the extent to which the investigator is initially able to make his samples identical in all but their air pollution exposures. Moreover, even if the samples are identical, the outside observer gets the strong impression that there exists great confusion about the criteria for experimental design, the physiological and metabolic responses that constitute excess health impacts, the validity of extrapolating from animals to humans, and the processes that generate any defined health impact [3].

As is well known, the multivariate regression procedures usually used by economists investigating the health effects of air pollution allow explicit discrimination between the effects of air pollution, the effects of other observed control factors, and the effects of unobserved, presumably random factors. Although the estimated health effects of pollution will be biased if some of the assumed random factors vary systematically with pollution, the continuous covariation between health states and pollution that the procedures permit does not force one to adopt the ambiguous notion of a human health effects threshold before research is even initiated [4]. Neither is the investigator put in the uncomfortable position of having to assign the residual ("excess" deaths or illnesses) to something particular such as air pollution.

The first attempt to investigate the health effects of air pollution at a national level without the resumption of a threshold was the pathbreaking effort of Lave and Seskin (1970). Using 114 U.S. metropolitan areas as units of analysis, they employed single equation, ordinary-least-squares methods to regress 1960 mortality rates linearly upon ambient concentrations of sulfates and particulate, and other plausible influences upon mortality. They tentatively concluded that statistically significant health effects of air pollution existed. This original study has inspired a substantial number of similar subsequent studies, including the culminating effort of Lave and Seskin (1977) [5]. Without exception, all have discerned a close and substantial inverse association between mortality rates and one or more air pollutants. Recently however, two studies have become available that should give considerable pause to those wishing to accept the Lave-Seskin, et al. findings.

Smith (1977), using data for 50 U.S. metropolitan areas in 1968-1969, applied versions of the Ramsey (1969) tests for specification error in the general linear model to 36 different single equation specifications. These specifications were similar, and often identical, to those greeted with the

most approval by the authors of the Lave-Seskin, et al. literature. None of the specifications could pass all of the Ramsey (1969) tests at the 10 percent level, although four passed all tests except that for non-normal errors.

The Ramsey (1969) tests are meant to be used to assess conformity with the basic assumptions for error structure of the classical linear model. They give no hint about events when attempts are made to correct for one or more of the specifications errors. In a recent paper, Crocker-Schulze, et al. (1979, pp. 24-71) use 1970 mortality data from 60 cities while trying to correct for potential omitted independent variable and simultaneous equation problems. Upon adding measures of medical care, cigarette consumption, and diet to the single equation Lave-Seskin, et al. specifications, they found no statistically significant effect of nitrogen dioxide, total suspended particulate, and sulfur dioxide upon the rate of total mortality[6]. Retaining the former variables, and accounting for the plausible simultaneity between health status and medical care, did nothing to improve the statistical significance of the three air pollution variables. On the presumption that these findings were sufficient to demonstrate the weakness of the Lave-Seskin type results, the authors did not go on to account for the obvious simultaneity between median age (or percentage over 65 years) and mortality incidence, income and mortality incidence, and several other plausible sources of simultaneity.

The results obtained by Smith (1977) and Crocker-Schulze, et al. (1979) cast doubt upon the robustness of the Lave-Seskin, et al. estimates, in spite of the no-threshold perspective embodied in these estimates. Nevertheless, before dismissing the hypothesis of an inverse relation between everyday air pollution levels and health states, it must be recognized that Lave- Seskin, et al, may have been asking more of their data than it was capable of giving'. Less than one in every 100 people dies in the U.S. each year. No biomedical authority asserts that air pollution is the dominant cause of the deaths that do occur. Many take the view that it is the direct cause of no more than a small fraction of these deaths, although they would agree that it may be quite important in intensifying predispositions toward mortality. However, the general properties of the underlying processes that encourage this predisposition are ill-understood. Thus, even with quite large samples, available estimation techniques and a <u>priori</u> knowledge may be inadequate for distinguishing the mortality effects of air pollution in a human population sample from a host of similar and plausible minor contributing factors.

The possible inadequacy of many available techniques for estimating the existence and/or magnitude of air pollutant-induced mortality applies with special force, given the data Lave-Seskin and their successors had to employ. Their work can be interpreted as an attempt at establishing the probability of

a representative individual currently residing in a representative region dying in a given year from a geographically representative level of air pollution occurring in a representative year. Since they had no information about the distribution of influential health factors, including air pollution, across the urban areas constituting their units of analysis, the identifying variabilities of their samples were perhaps drastically reduced. [87] When this relatively low variability of the samples is coupled with what are probably substantial measurement errors in the air pollution variables, the baggage of additional explanatory variables and more sophisticated estimation techniques to correct for specification error that the data are able to carry must be rather light. The attempted corrections may serve only to misinform. Furthermore, that which is being corrected may be only an apparition since, as Crocker (1975, pp. 350-351) demonstrates, the measure of (the probability of) death, employing some group of individuals as the fundamental unit of observation, can differ from one group to another; there could be as many unique measures employed as there are groups.

The preceding remarks lead us to three conclusions. First, given the biomedical and economic subleties inherent in comprehending the etiologies of air pollution-induced mortality and morbidity, the estimates obtained from aggregated data used in the great bulk of extant studies are unlikely ever to be sufficiently compelling to establish a consensus. Only the use of actual individuals as fundamental units of observation is likely to provide enough strength in the data base to carry the requisite statistical burdens. Second, the statistical burdens that have to be carried might be considerably lightened if research concentrates on morbidity rather than mortality. The frequency, and most likely the identifying variability, of the former is greater by a factor of fifteen or twenty. Finally, because one's health status is influenced by the choices one makes about lifestyles, environmental and occupational exposures to possible toxics, and other health-influencing factors, economics can provide a priori hypotheses and an analytical framework to lend additional structure to epidemiological investigations. The relationships with which observed real world outcomes are consistent can, therefore, be further narrowed.

A CRITICAL REVIEW OF OUR WORK

Crocker-Schulze, et al., (1979) embodies both mortality and morbidity studies. The mortality study had the essentially negative purpose of empirically demonstrating that the estimates derived in Lave-Seskin type studies are not at all robust. The morbidity study had the more positive purpose of investigating air pollution and human health status with a data set better able to bear added statistical burdens and to accept hypothesis testing about the impact of man's free will upon health status. In this section, we briefly

discuss several entirely correct ways in which the morbidity study is susceptible to injury. Strangely, although the study has been carefully pursued by many interested parties, few have hit it where their thrusts could not even begin to be countered without additional work on our part. Here, we present some of those thrusts.

Depending almost entirely upon ordinary-least-squares (OLS), the morbidity study estimated the effect of air pollution upon self-reported health status measured as length of time chronically ill and annual frequency of acute illnesses. Expressions linear in the original variables were estimated for several 400 person samples independently drawn from all household heads in the Panel Survey of Income Dynamics (PSID) [Survey Research Center (1972)] who had always lived in one state. Although some attention was devoted to NO_2, air pollution was generally measured as the annual 24-hour geometric mean of SO_2 and/or TSP in the head's county of residence for the year (196775) from which the sample was drawn. In addition to air pollution, measures of the intensity of the head's illness, his biological and social endowments, life-style, and work, home, and outdoor environments were, when available, included as explanatory variables. Air pollution contributed positively and significantly to both chronic and acute illnesses in the majority of the unpartitioned samples. Upon combining these dose-response estimates with a simple recursive labor supply formulation, the economic impact of air pollution-induced chronic illness upon labor productivity was estimated to exceed that of air pollution-induced acute illness by nearly a factor of 20.

These results encouraged us to proceed further, particularly with respect to investigating air pollution-induced chronic illness. The obvious initial further step was to correct some of the outstanding technical problems in our treatment of the dose-response functions estimated from the PSID data. [97] These problems fall into three general categories: (1) the definition of self reported health status; (2) the factors used to explain self-reported health status; and (3) the algorithm used to estimate self-reported health status.

The PSID data on the chronic illness health status of household heads consists only of responses to four questions stated in the following order:

1. Do you have a physical or nervous condition that limits the type of work you can do or the amount of work that you can do?

2. How much does it limit your work?

3. How long have you been limited in this way by your health?

4. Is it getting better, worse, or staying about the same?

In the case of the first question, persons were asked for a yes or no answer, while for the remaining three questions the response called for was categorical. The response to question #3 was used as the dependent variable in our earlier analysis. However, the responses to this question were recorded categorically with the uppermost category being bounded only by age. Moreover, this response was conditional upon the response to question #1 and possibly question #2. For these reasons, interpretation of the earlier chronic illness dose-response estimates required a string of assumptions that may or may not have been important to stated results. In any case, in order to assess the validity of the earlier results, it is preferable to remove any clouding that the assumptions may have introduced. The response to question #1 is unambiguous.

Even though the response to question #1 is unambiguous in terms of self reported health status, it need not represent the respondent's clinical health status. More specifically, individuals may not be alike in the way they determine whether or not they are chronically ill. Economic factors including type of job, access to disability benefits, and other measures of the opportunity costs of not working may be important to this determination. For example, consider two persons who are alike in every respect other than their hourly wage. The person with the lower of the two wage rates will have a lower opportunity cost of not working. He may be perfectly healthy but desire to work fewer hours and use illness as an excuse, or he may actually be sick more often than his higher income counterpart because he does not find it economically advantageous to be as healthy.

The preceding suggests that our earlier estimated chronic illness dose-response expressions might be biased because economic determinants of self reported health status were omitted. In addition to these economic determinants, other, more traditional life-style, biological endowment, medical care, and environmental determinants were omitted or imperfectly measured. For example, the earlier estimates included no information on job accident rates, and used cigarette expenditures as an index of cigarette consumption. These variable exclusions and imperfectly measured explanatory variables can bias the estimated contribution of air pollution to self-reported health status.

Finally, given the chronic illness health status variable employed in our earlier work, the use of an OLS estimation procedure could have been inappropriate for two reasons. First, self-reported health status might have been determined jointly with some explanatory variables (e.g., leisure exercise, cigarette smoking, and medical care) that were also choice variables. OLS estimates of the chronic illness dose-response expression would

then be biased and inconsistent. Second, the health status variable was recorded in a categorical rather than in a continuous fashion. This means that hetero-skedasticity could be present in the OLS-estimated chronic illness dose-response expressions with a consequent introduction of biases in the standard errors of the air pollution coefficients. As McKelvey and Zavoina (1975) show, the use of OLS procedures with categorical dependent variables can cause the relative impacts of certain variables to be severly underestimated.

SOME NEW, BUT LIMITED RESULTS

In this section, we present some new results which, insofar as available data allow, correct partially or wholly for the technical problems raised in the previous section. The outstanding failing of these new results is that we do not construct an explicit analytical model to account for the economic determinants of self-reported health status. Instead, we do no more than introduce explanatory variables such as family assets and union membership that would plausibly have a role to play in expressions derived from any analytical model dealing with the effect of the opportunity costs of not working upon perceived own health status.

Table 1 lists the variables we employ. Alcohol expenditures, numbers of daily cigarettes smoked, free access to medical care, physician population, carcinogenic potential in the workplace, precipitation, workplace job accident rate, current transfer income, and union membership all represent variables that did not appear in our previous chronic illness dose-response expressions. Separate structural expressions are estimated for numbers of daily cigarettes smoked, whether or not the individual has medical insurance, and whether or not he participates in strenuous leisure exercise on the presumption that they are jointly determined with health status. To account for plausible nonlinearities with respect to the impact of age and food expenditures on health status, squared, as well as original, values are entered for these variables.

In view of the categorical nature and the simultaneity of the dependent variable, the estimation technique selected was the two-stage limited dependent variables (2SLDV) approach suggested by Nelson and Olson (1978). More specifically, the estimation procedure these authors propose is to:

(i) Estimate the reduced form of the structural systemby applying an appropriate maximum likelihood technique to each.

(ii) Form instruments from the “predicted” values of the

TABLE 4.1

COMPLETE VARIABLE DEFINITIONS

Self-Reported Health Status Variables

DSAB - Limitation on work = 1; otherwise = 0

LDSA - Disabled for ≤ 2 years = 1; 2-4 years = 2; 5-7 years = 3; ≥ 8 years = 4; otherwise = 0.

Biological and Social Endowment Variables

AGE - Age in years.

EDUC - Completed 6-8 grades = 2; 9-11 = 3; 12 grades = 4; 12grades plus non-acedemic training = 5; college, no degree=6; college degree = 7; advanced or professional degree = 8; otherwise = 1.

FMSZ - Family size in number of persons in housing unit.

POOR - Stated that parents were poor ".. ..when you were growingup..." =1; otherwise = 0.

SEX - Male = 1; Female = 0.

Lifestyle Variables

ALKY - Annual alcohol expenditures X 10^2 per adult family member.

CIGN - Number of daily cigarette packs smoked per adult family member. This variable was calculated by dividing the PSID data on 1970 cigarette expenditures by the 1970 retail price of a pack of cigarettes in the 1970 state of residence. Retail price data was taken from Tobacco Tax Council, Inc. (1978, pp. 67-69).

FOOD - Family food consumption relative to food needs standard in percent. Consumption refers to food expenditures in dollars and includes amounts spent in the home, school, work, and restaurants, as well as the amount saved in dollars by eating at work or school, raising, canning, or freezing food, using food stamps, and receiving free food. The food needs standard is in dollars and is based on USDA Low Cost Plan estimates of weekly food costs as published in the March 1967 issue of the Family Economics Review. The standard itself-is calculated by multiplying the aforementioned weekly food needs by 52 and making a series of adjustments according to family size.

LEXR - Indication that dominant leisure-time activities involves strenuous exercise = 1; otherwise = 0. Strenuous activities were said to include fishing, bowling, tennis, camping, travel , hunting, dancing, motorcycling, etc.

Health Care Variables

HVET - Free access to medical care as a veteran or through medicaid = 1; otherwise = 0.

INSR - Has hospital or medical insurance = 1; otherwise = 0.

PHYS - Physicians per 10,000 population in county of residence on July 1, 1975. This data was obtained from U.S. Bureau of the Census (1978, Table 2).

Environmental Variables

CANX - An index.of workplace "carcinogenic potential" by two-digit SIC code as presented in Hickey and Kearney (1977) and determined by dividing their Table 8 by their Table 7. We are aware that these authors insist that "... the magnitude of the derived carcinogenic potential is not suitable for any health hazard inference" (p. iii).

COLD - Mean annual January temperature in the 1970 county of residence in F" X 10. This data is from U.S. Bureau of the Census (.1978, Table 4).

PRCP - Mean annual precipitation in inches X 10^2 in the 1970 county of residence. This data is from U.S. Bureau of the Census (1978, Table 4).

JACCR - Number of disabling work injuries in 1970 by 2 and 3-digit SIC code for each million employee hours worked. The data is from Table 163 of Bureau of Labor Statistics (1972).

SULM - Annual 24-hour geometric mean sulfur dioxide micrograms per cubic meter as measured by the Gas Bubbler Pararosaniline-Sulfuric Acid Method. The data were obtained from the annual USEPA publication, Air Quality Data - Annual Statistics, and refer to a monitoring station in the 1970 county of residence.

TSPM - Annual 24-hour geometric mean total suspended particulate in micrograms per cubic meter as measured by the Hi-Vol Gravimetric Method. The data were obtained from the annual USEPA publication, Air Quality Data - Annual Statistics, and refer to a monitoring station in the 1970 county of residence.

Pecuniary Variables

ASSETS - Sum of 1970 income in dollars X 10^2 from social security, retirement pay, pensions, annuities, dividends, interest, and rent.

UNION - Member of a labor union = 1; otherwise = O.

dependent variables using the observations from the sample on the exogenous variables together with the estimated reduced from coefficients obtained in the first step.

(iii) Replace the jointly dependent variables on the righthand side of the equations in the structural system with their instruments constructed in the second step.

(iv) Estimate the resulting relations by an appropriate maximum likelihood method.

As can be easily seen, this estimation procedure applied to a system of simultaneous equations is just two-stage least squares in the case where all jointly dependent variables are continuous over the entire real line. However, the approach of Nelson and Olson (1978) takes account of the fact that some dependent variables, particularly the DSAB variable of interest here, do not exhibit this type of behavior. They therefore suggest that an appropriate limited dependent variable technique be used in the estimation of both the reduced form and the structural form of the model. In this case, since DSAB is defined to take on only the values of zero or one, the probit model would appear to be the most appropriate of the alternative limited dependent variable methods.

The procedures outlined above were applied to a sample of 309 individual household heads drawn from the 1970 calendar year of the PSID sample. All individuals had always resided in the 1970 state of residence. We are, thus, able to control partially for the air pollution exposure history of the individual, given that relative 1970 pollution concentrations across residential locations are similar to the history of relative concentrations, The year 1970 was selected for detailed empirical analysis because the chronic illness dose-response expressions estimated for this year in Crocker-Schulze, et al. (pp. 105-109) were considered to be the best representatives of all the expressions for assorted years estimated by ordinary-least-squares from the PSID data.

The 309 individuals of the sample represent all individuals in the 1970 PSID calendar year data for whom we were able to obtain observations on each explanatory variable, including total suspended particulate and sulfur dioxide. It should be noted that this sample is unlikely to correspond to a random sample of the U.S. population. If anything, as a glance at the arithmetic mean values of the explanatory variables presented in Table 2 shows, the sample appears to include a somewhat disproportionately high number of female household heads, “poor” childhood backgrounds, and relatively low pecuniary values of family assets. For our present purposes, of course, a

random sample is unnecessary, given that the sample was not selected on the basis of whether or not the individual reported he suffered from a chronic illness.

The results of estimating the augmented (relative to our previous work) chronic illness dose-response' expression by the multivariate Probit estimator are reported in the last two columns of Table 2. As Poirier and Melino (1978) have shown, the coefficients are proportional to the change in the probability that an individual will report being chronically ill for a one unit change in the explanatory variable. Thus, for example, a male, is nearly twice as likely to report being chronically ill as is a female. Our use of the Probit estimater presumes that each individual has a threshold level of the explanatory variable below which he will not view himself as being made chronically ill. However, the estimator also presumes that there exists a transformation causing these threshold values to be normally distributed over our sample and, therefore, that there exist some individuals for whom even . minor levels of air pollution will cause them to report being chronically ill. The constant term is simply a shifter.

With the exceptions of CIGN, LEXR, and POOR, the signs of all coefficients coincide with a priori expectations. The combinations of signs for the AGE variables and the FOOD variables are consistent with increased likelihoods of reporting chronic illness at the extremes of age and diet adequacy with a reduced likelihood in the middle ranges. Increases in alcohol consumption, exposures to carcinogenic substances, accident risks in the workplace, physicians to originate or confirm the individual's self-diagnosis, and air pollution in the form of sulfur dioxide all serve to increase the chances of self-reported chronic illness. The coefficients of CANX and JACCR are probably biased downward, since they refer only to the current workplace, rather than to the individual's workplace history. On the other hand, consistent with the work of Tromp (1962) and others, high precipitation and low midwinter temperatures are less likely to make the individual feel chronically ill. Those variables such as ASSETS and UNION, representing factors thought to reduce the opportunity costs of feeling chronically ill, all contribute positively to the probability of reporting chronic illness. Similarly, more education and larger family size, variables which capture factors tending to increase the opportunity costs of feeling chronically ill, each have negative signs attached. Since people who are veterans and have medical insurance face lower marginal prices for medical care, they can be expected to consume more medical care thereby reduce the frequency of their chronic illnesses. The negative signs attached to HVET and INSR are consistent with this interpretation. Note that the coefficient attached to the latter variable is estimated from a system that accounts for the simultaneity between the likelihood of possessing medical insurance and the

TABLE 4.2

MAXIMUM LIKELIHOOD ESTIMATES OF SELF-REPORTED CHRONIC ILLNESS (DSAB)

Variable		Mean	Coefficient	Standard Error
AGE		39.36	0.084	0.054
$(AGE)^* \times 10^{-1}$		177.00	-0.776	0.582
ALDY		1.11	0.169	0.100
ASSETS		2.68	0.001	0.001
CANX		18.77	0.006	0.021
$\widehat{CIGN}$	$\{CIGN =$	1.73		
	$\{\widehat{CIGN} =$	0.64	-0.527	0.190
COLD		37.86	-0.025	0.015
EDUC		3.76	-0.087	0.162
FMSZ		3.22	-0.005	0.056
FOOD		1.80	-0.499	0.470
$(FOOD)^2$		3.90	0.089	0.095
HVET		0.19	-0.472	0.400
$\widehat{INSR}$	$\{INSR =$	0.72		
	$\{\widehat{INSR} =$	0.80	-1.223	0.490
JACCR		33.17	0.003	0.005
$\widehat{LEXR}$	$\{LEXR =$	0.18		
	$\{\widehat{LEXR} =$	-1.13	0.115	0.454
PHYS		24.08	0.007	0.010
POOR		0.52	-0.503	0.290
PRCP		39.77	-0.043	0.017
SEX		0.57	0.927	0.556
SULM		18.37	0.011	0.010
UN 10N		0.19	0.422	0.398
Constant			1.090	1.807

(-2.0) times log of likelihood ratio 85.609; statistically significant at the one percent for the χ^2 distribution with 21 degrees of freedom.

Observations at Unity 77
Observations at Zero 232

NOTE: No levels of significance are indicated because the asymptotic properties of the standard errors for this sample are not known. A simulation experiment with the simultaneous probit estimator suggested to Nelson and Olson (1978, p. 702) that its standard errors could be biased upward by as much as a factor of 1.6.

presence of chronic illness. Note also, however, that the results for these variables explaining the "demand" for chronic illness have not been derived from an explicit analytical model. The above interpretation may therefore be unwarranted.

Interpretations for the signs of CIGN, LEXR, and PCOR are less readily provided. It is possible that no one of these variables is a reasonable measure of the effect we were trying to capture. For example, CIGN represents the estimated number of current cigarettes smoked per adult family member. There is no obvious connection between this measure and the smoking history of the individual whose health status is being inspected. It is, of course, possible that those who are already chronically ill increase their smoking because of the greater utility it might then afford. As for LEXR, it appears from its estimated mean value that the expression used to calculate it did not perform very well. In addition, the perception of what constitutes strenuous exercise can differ across individuals. Again, strenuous exercise might yield greater utility for those who are already chronically ill, so that they are more likely to participate in it than are healthy individuals. Similarly, the current perception of whether one's parents were poor may be more a measure of one's current real income status relative to the former status of one's parents rather than an absolute measure of the latter's former status. Thus, extending the Dusenberry (1949) hypothesis to an intergenerational context, it might be that greater relative current real income may engender a sense of security reducing the opportunity costs of being chronically ill. Alternatively, the explanation for the unexpected negative sign might simply be that a selection process operated in the past to eliminate those who were less well genetically endowed and who also had poor childhoods.

A rank-ordering of the explanatory variables from the most to the least statistically significant results in the following: CIGN, INSR, PROP, POOR, ALKY, SEX, COLD, AGE, $(AGE)^2$, HVET, FOOD, UNION, SUM, ASSETS, $(FOOD)^2$, PHYS, JACCR, EDUC, CANX, LEXR, AND FMSZ. Thus, at least for the sample represented in Table 2, air pollution, as measured by annual 24-hour geometric mean sulfur dioxide, is less robust statistically than the climate variables but more robust than the measures of occupational hazards. However, as indicated in the table, SULM would appear to be statistically insignificant at conventional levels. This general conclusion holds when another air pollution variable, annual 24-hour geometric mean suspended particulate, replaces the measure of sulfur dioxide used in Table 2. Upon doing this, a coefficient of 0.006 with a standard error of 0.007 is obtained. Given that the standard errors of the simultaneous probit estimator are thought to be biased upward (perhaps by as much as 1.6 according to Nelson and Olson (1978, p. 702), the actual effect of air pollution on self-reported health status may be more significant than our results indicate. Nevertheless, even if the standard error on the air

pollution coefficients are in fact biased upward by a factor of 1.6, the statistical significance of these coefficients remains questionable.

In order to provide another basis for comparison with Crocker-Schulze, et al. (1979), we substituted the measure used for the length of chronic illness (LDSA) in our earlier work for the dependent variable in Table 2. The system was estimated by the two-limit simultaneous probit technique employed in Nelson and Olson (1978). Again, the results obtained were not inconsistent with our previous OLS estimates. In fact, the magnitudes of the air pollution coefficients were almost twice those obtained in the OLS results. However, as Poirier and Melino (1978) demonstrate, the coefficients of an explanatory variable in a truncated regression procedure such as probit is proportional to, but not equal to, the partial derivative of the conditional mean of the dependent variable with respect to a one unit change in an explanatory variable. This factor of proportionality, which is identical for each coefficient in a regression, can be determined when the variance of the untruncated variable is known. For the PSID data set, this variance is unknown.

WHITHER FROM HERE

The motivation for this paper, as well as our previous work in the area, originated in our convictions that economic analysis and its empirical techniques could contribute to the resolution of certain recurring puzzles in studies of the incidence and severity of diseases in human populations, particularly the epidemiology of air pollution. We have viewed human health status as a decision variable and have therefore been able to employ economic theory as a means of providing more a priori structure for the analysis of epidemiological data. Considering only the empirical results reported in the previous section, it seems we have not yet provided enough information on structure for resolution. We have by no means, however, exploited all the conceivable economic-behavioral structural relations from which restrictions might be obtained.

One might introduce more statistical information by quasi-replication of the structures already estimated; that is, we could pull additional samples from the PSID data set and estimate for each of those samples the same two structures already discussed. This strategy has been used [Crocker-Schulze, et al. (1979)] in an earlier substantially less rigorous treatment of the same data.

Alternatively while retaining the structure that economic analysis and epidemiology provide, we can draw upon knowledge in biophysics, biochemistry, and bioenergetics to a much greater degree than previous studies in air

pollution epidemiology appear to have done. In a manner consistent with human capital theory, as some existing work has in fact already done [e.g., Cropper (1977) and Crocker-Schulze, et al. (1979)]. The individual might be construed as having an initial health endowment that, due to natural aging, depreciates exogenously over time. However, by his decisions about life-style and his occupational and environmental exposures, he can either slow or accelerate this natural depreciation. An integral part of these human capital treatments has been the representation of a production function in implicit form where some crude measure of health status is determined by rather arbitrary assortments of the aforementioned collection of life-style, occupational, and environmental variables. We suggest, at least insofar as empirical treatments are concerned, that one can specify this production function in much more detail while retaining the human capital framework for the individual's decision problem.

As an alternative to traditional toxicological research emphasis upon metabolities and metabolic pathways, the Second Task Force for Research Planning in Environmental Health Science (1977, Chapter 14) recommends that more effort be devoted to building upon existing knowledge of the structure and function of particular organ systems such as the respiratory and cardiovascular systems. Contrary to most of the arcane (to an economist) basic research on the fundamental chemical processes at work in various metabolic pathways, much of the work on the determinants of the individual's research of organ function appears to be readily translatable into mere displays of the fact that within limits the same quality of some simple measure of the health status of the organ system, such as the ventilation capacity of the lung, can be obtained from various combinations of inputs[10]. In many cases, the responses of the health indicator of the organ system to various stresses follow well- known physical laws having specific functional forms and even particular values attached to coefficients.[11].

When writing down the individual's decision problem with respect to health status, we may be able to structure the problem more tightly by building the aforementioned information on organ system responses directly into the constraint set. Rather than having an implicit production function in which the value of a "self-reported, highly aggregated measure of health status (e.g., whether or not the individual is chronically ill) is explained by a collection of intuitively reasonable variables, one can employ a description that precisely maps a limited and well-defined set of major influential factors into a continuous scaler measure of the health of an organ system.

SUMMARY AND CONCLUSIONS

The preceding pages are not without technical sin. In particular, with

out rigorously explaining from whence they cone, we have introduced variables that are supposed to represent the opportunity costs of reporting or failing to report ones self chronically ill. Otherwise, however, by employing a more robust estimation procedure, by redefining the chronic illness variable, and by introducing better measures of cigarette smoking, hazards and toxic exposures in the workplace, medical care, and climate, we have responded to several well-founded. criticisms of the morbidity results in Crocker-Schulze, et al. (1979). On the basis of those new tests, we see no reason to alter our previous interpretation of the effect of air pollution upon self-reported chronic illness.

REFERENCES

1 In accordance with the eloquent argument of Calabresi and Bobbit (1978), one might attribute the dominance of this perspective in public policy settings to the fictions erected by societies to segment markets that would otherwise require explicit judgments about the relative worths of individuals' lives. Calabresi and Bobbit (1978) argue that these fictions seine to soften intolerable societal stresses. The purpose they serve in a scientific setting is not obvious.

2 Alternatively, the laboratory studies try to specify the intervening processes causing an observed health effect.

3 Apart from these issues, the practice of applying laboratory results to everyday human environments is questionable. As Anderson and Crocker (1971, p. 146) note, so as to remove all sources of stress other than air pollution, all other factors influencing health in the laboratory tend to be set at biologically optimal levels. Given that these biologically optimal levels exceed those found in everyday environments, it follows from the law of variable proportions that air pollution-induced health effects in the laboratory will exceed those found in everyday environments.

4 It should be noted that many biomedical authorities strongly dispute the biological existence and the policy relevance of thresholds for most environmental contaminants. Authors such as Epstein (1974), Goldsmith and Friberg (1977) argue that any positive amount of pollution induces ill-health effects for some individuals and increases the probability of ill-health for everyone exposed.

5 Among the more notable examples are: McDonald and Schwing (1973); Liu and Yu (1976); Mendelssohn and Orcutt (1979); Gregor (1977) and Koshal and Koshal (1973).

6 However, particulate was statistically significant in an expression explaining pneumonia and influenza related deaths. Sulfur dioxide was

statistically significant in an expression for deaths attributed to early infant diseases. Nitrogen dioxide would have been statistically significant in heart disease if a slightly less severe level of acceptance had been adopted.

7 In order to get the data to "give" more, the authors of the Lave-Seskin type work have' usually tested with the same data set several different functional forms and combinations of explanatory variables. The objective frequently seems to have been the maximization of certain summary statistics (e.g., the coefficient of determination) having no basis in any *a priori* hypothesis. We are unaware that the pretest or selection procedures surveyed in Wallace (1977) and Judge, et al. (1980, Chap. II) have ever been employed during these manipulations. If these procedures are not employed, the properties of the classical least squares estimators these authors typically use can be substantially altered; that is, the customary interpretations cannot be attached to estimated coefficients and standard errors.

8 Ambient pollution concentrations for a single year at single (usually downtown) sites served as proxies for the lifetime exposure histories of entire regional populations. For a succinct treatment of the trade-off between corrections for specification error and identifying variability when measurement error is present in an independent variable of interest, see Griliches (1977, pp. 12-13). The addition of imperfectly measured explanatory variables to the expression being estimated will bias downward the coefficients of the air pollution variables.

9 For now, we much prefer to leave accounting issues about what the estimate mean in terms of national economic impacts to more adventuresome types.

10 See Kao (1972, Chap. 111 and IV) for readily understood treatments of the lung as a mechanical pump and as a gas exchanger.

11 Many of these responses have been established in animal rather than human studies. The validity of extrapolating results from the former to the latter is a major source of controversy in biomedical studies of pollution effects upon organ systems.

BIBLIOGRAPHY

Anderson, R.J., Jr., and T.D. Crocker, "The Economics of Air Pollution: A Literature Assessment," in P.B. Downing, cd., Air Pollution and the Social Sciences, New York: Praeger Publishers (1971), 133-166.

Bouhuys, A., G.J. Beck, and J.B. Schoenberg, "DO Present Levels of Air Pollution Outdoors Affect Respiratory Health?" Nature 276(Nov. 30, 1978), 466-471.

Bureau of Labor Statistics, Handbook of Labor Statistics, 1972, Bull. 1735, U.S. Department of Labor, Washington, D.C.: USGPO (1972).

Calabresi, G., and D. Bobbitt, Tragic Choices, New York: W.W. Norton (1978).

Crocker, T.D., "Cost Benefit Analysis of Cost-Benefit Analysis," in H.M. Pesken and E.P. Seskin, eds., Cost-Benefit Analysis and Water Pollution Policy, Washington, D.C.: The Urban Institute (1975), 341-360.

Crocker, T.D., W. Schulze, S. Ben-David, and A.V. Kneese, Experiments in Air Pollution Epidemiology, Washington, D.C.: USEPA Publication No. 60015-79-001a (1979).

Cropper, M.L., "Health, Investment in Health, and Occupational Choice," Journal of Political Economy 85(Dec. 1977), 1273-1294.

Duesenberry, J.S., Income, Saving, and the Theory of Consumer Behavior, Cambridge, Mass.: Harvard University Press (1949).

Engel, G.L., "The Need for a New Medical Model: A Challenge for Biomedicine," Science 195(Jan. 22, 1977), 129-136.

Epstein, S. "Environmental Determinants of Human Cancer," Cancer Research, 34(Oct. 1974), 2425-2435.

Goldsmith, J.R., and L.T. Fribert, "Effects of Air Pollution on Human Health,"

in A.C. Stem, cd., The Effects of Air Pollution, 3rd cd., New York: Academic Press (1977).

Gregor, J.J., Intra-Urban Mortality and Air Quality, Corvallis, Ore.: USEPA Pulbication No. 60015-77-009 (1977).

Griliches, A., "Estïmating the Returns to Schooling: Some Econometric Problems," Econometrics, 45(Jan. 1977), 1-21.

Hickey, J.L.S., and J.J. Kearney, Engineering Control Research and Development Plan for Carcinogenic Materials, Cincinnati, Ohio: U.S. Public Health Service under Contract No. 210-76-0147 (Sept. 1977).

Judge, G.G., W.E. Griffiths, R.C. Hill, and T. Lee, The Theory and Practice of Econometrics, New York: John Wiley and Sons (1980).

Kao, F.F., An Introduction to Respiratory Physiology, Amsterdam: Excerpts Medics (1972).

Koshal, R.K., and M. Koshal, "Environments and Urban Mortality: An Econometric Approach," Environmental Pollution 4(June 1973), 247-259.

Lave, L.B., and E.P. Seskin, "Air Pollution and Human Health," Science 169(August 21, 1970), 723-733.

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

Liu, B., and E. Yu, Physical and Economic Damage Functions for Air Pollutants by Receptor, Corvallis, Ore.: USEPA Publications No. 60015-76-011 (1976).

McDonald, G.C., and R.C. Schwing, "Instabilities of Regression Estimates Relating Air Pollution to Mortality," Technometrics 15(1973), 463-481.

McKelvey, R.D., and W. Zavoina, "A Statistical Model for the Analysis of Ordinal Level Dependent Variables," Journal of Mathematical Sociology 4(1975), 103-120.

Mendelssohn, R., and G. Orcutt, "An Empirical Analysis of Air Pollution Dose-Response Curves," Journal of Environmental Economics and Management 6(1979), 85-106.

Morris, S.C., M.A. Shapiro, and J.H. Wailer, "Adult Mortality in Two Communities with Widely Different Air Pollution Levels," Archives of

Environmental Health, 31(1976), 248-254.

Nelson, F., and L. Olson, "Specification and Estimation of a Simultaneous-Equation Model with Limited Dependent Variables," International Economic Review, 19(October 1978), 695-709.

Poirier, D.J. and A. Melino, "A Note on the Interpretation of Regression Coefficients within a Class of Truncated Distributions," Econometrics, 46(September 1978), 1207-1209.

Ramsey, J.B., "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis," Journal of the Royal Statistical Society Series B, 31(1969), 350-371.

Second Task Force for Research Planning in Environmental Health Science, Human Health and the Environment: Some Research Needs, Washington, D.C.: USDHEW Publication No. NIH77-1277 (1977).

Smith. V.K., The Economic Consequences of Air Pollution, Cambridge, Mass.: Ballinger Publishing Co. (1977).

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

Tobacco Tax Council, Inc., The Tax Burden on Tobacco, Washington, D.C.: Tobacco Tax Council, Inc. (1978).

Tromp, S.W., Medical Biometeorology, Amsterdam: Elsevier Publishing Co. (1962) .

U.S. Bureau of the Census, County and City Data Book, 1977, Washington, D.C.: U.S. Government Printing Office (1978).

Wallace, T.D., "Pretest Estimation in Regression: A Survey," American Journal of Agricultural Economics, 50(August 1977) 431-443.

Chapter V
MEASURING THE BENEFITS FROM REDUCED ACUTE MORBIDITY

INTRODUCTION

The predominant view in economics is that individuals are unaware of the health effects of air pollution and therefore do not take them into account in making decisions (Lave 1972). Given this view, the appropriate way to measure the morbidity benefits of a reduction in pollution is to estimate a damage function and then assign a dollar value to the predicted decrease in illness. This, together with any reduction in medical costs, is what an individual would pay for a decrease in pollution if he treated his health as exogenous.

Unfortunately, this approach is inconsistent with the view, widely held in health economics, that individuals can affect the time they spend ill by investing in preventive health care. Support for this view is provided by Michael Grossman (1972a, 1972b, and 1975) whose work indicates that individuals diet, exercise and purchase medical services to build up resistance to illness. These findings suggest that if persons in polluted areas perceive their resistance to illness decreasing they will try to compensate by exercising more, smoking less or getting more sleep. Conversely, an improvement in air quality should lead to a decrease in preventive health care, and the value of this must be added to the benefits of pollution control.

Human capital theory thus implies that the damage function approach, by ignoring the value of preventive health care, understates willingness to pay for a change in air quality. This conclusion, it should be emphasized, does not assume that individuals know precisely the medical effects of air pollution. All that is necessary for a person to try and compensate for the effects of pollution is that he feels worse when pollution increases.

This paper presents a simple model of preventive health care, similar to that of Grossman (1972a, 1972b), and uses the model to define what a person would pay for a change in air quality. The model assumes that one can build up resistance to acute illness by increasing his stock of health capital; however, health capital decays at a rate which depends on air pollution. For

acute illness, willingness to pay as derived from the model, is greater than the benefit estimate computed using the damage function approach. To illustrate the size of this discrepancy estimates of willingness to pay are computed using data from the Michigan Panel Study of Income Dynamics.

A MODEL OF INVESTMENT IN HEALTH

The essence of the human capital approach to health is that each individual is endowed with a stock of health capital, H, which measures his resistance to illness. This stock can be increased by combining time, TH_t, with purchased goods, M_t, to produce investment in health,

$$I_t = TH_t^{1-\zeta} M_t^{\zeta} E_{1t}^{\xi_1} \cdots E_{nt}^{\xi_n}. \qquad (1)$$

Outputs of equation (1) include exercise, rest and nourishment. These will be affected by factors such as the individual's knowledge of health, or the presence of a chronic disease ($E_{it}, \ldots, E_{nt}$ in equation (1)).

For simplicity suppose that investment in health exhibits constant returns to scale so that the marginal cost of investment is constant and independent of I_t. This is reflected in equation (2) which gives the marginal cost of investment, π_t, as a function of the price of purchased goods, PM_t, and wage, W_t,

$$\pi_t = W_t^{1-\zeta} PM_t^{\zeta} E_{1t}^{-\xi} \cdots E_{nt}^{-\xi_n}. \qquad (2)$$

Investment in health increases the individual's health stock, H_t, according to equation (3),

$$dH_t/dt = I_t - \delta_t H_t, \qquad (3)$$

Health capital also deteriorates at the proportional rate δ_t since resistance to illness would decline if no investments were made in health.

The main motive for investing in health is that health capital affects time spent ill, TL_t. For empirical work it is most appropriate to assume a threshold relationship between health capital and illness since a large number of persons (half of the Panel Study sample) report zero days of illness each year. A discontinuous relationship between H_t and TL_t, however, makes the solution to the individual's choice problem difficult. We therefore assume that the individual views the log of illness as a decreasing function of the log of health capital.

$$\ln TL_t = y - \alpha \ln H_t, \qquad \alpha > 0. \tag{4}$$

This implies that time spent ill can be made arbitrarily small, although not zero.

Equations (3) and (4) suggest that the model, while appropriate for acute illness, should not be applied to chronic illness. In (4) a reduction in the health stock increases time spent ill; however, being ill in one instant does not reduce the stock of health capital in the next. This is reasonable only if TL_t refers to acute illnesses such as colds and the flu.

To simplify the model and facilitate estimation of willingness to pay (4) is assumed to be the only motive for investing in health. This reduces health to a pure investment good and implies that the only effect of health on utility is through the budget constraint.

In this case the decision to invest in health can be separated from the decision to purchase other goods. First, a path of investment in health is chosen to maximize R, the present value of full income net of the cost of investment, then utility is maximized, given R. In the present model full income is the market value of the individual's healthy time. If Ω is the total time available at t then $h_t = \Omega - TL_t$ is the amount of healthy time available. The present value of full income net of the cost of investing in health may therefore be written

$$\int_0^T [(W_t h_t - \pi_t I_t) e^{-rt} dt, \tag{5}$$

where T is length of life. The individual's problem is to choose the path of investment which maximizes (5) subject to (3) and (4).

When the marginal cost of investment is constant the solution to this problem is simple: at each instant the individual chooses an optimal level of resistance, H_t^*, and then determines the amount to invest in health from (3).[17] The optimal health stock is determined by equating the value of the marginal product of health capital, $W_t \partial h_t / \partial H_t$, to its supply price,

$$W_t \frac{\partial h_t}{\partial H_t} = \pi_t \left(r + \delta - \frac{d\pi_t}{dt} \frac{1}{\pi_t}\right). \tag{6}$$

The latter consists of three parts: the interest foregone by investing π_t in

health rather than at the rate r, the depreciation cost, $\pi_t \delta_t$, since each unit of health immediately declines by an amount δ_t, and a capital gain which accrues if the cost of investment is changing! If π_t is rising at approximately the rate of interest then the right-hand-side of (6) reduces to $\pi_t \delta_t$.

Substituting from (4) the optimal health stock may be written

$$\ln H_t^* = \frac{1}{1+\alpha} (\beta + \ln W_t - \ln \pi_t - \ln \delta_t), \quad \beta = \gamma + \ln a, \tag{7}$$

while time spent ill is given by

$$\ln TL_t^* = y - \frac{\alpha}{1+\alpha} (\beta + \ln W_t - \ln \pi_t - \ln \delta_t). \tag{8}$$

There are several ways that pollution could enter this model. The observation that individuals are ill more often in polluted environments could mean that pollution enters the equation for time spent ill, (4), with a positive coefficient. This, however, implies that two individuals with the same health stock are not really equally healthy. Instead, it seems preferable to assume that pollution physically alters the state of a person's health. [2/] This can be accomplished by making the rate of decay of health capital a function of air pollution, P_t,

$$\delta_t = \delta \cdot e^{\delta t} P_t^{\psi} S_t^{\phi} . \tag{9}$$

Equation (9) also implies that the rate of decay of health varies with age and with other factors, St, such as stress or pollution on the job. [3/]

Adding equation (9) to the model means that it is more costly to build up resistance to illness in polluted environments, hence individuals in polluted areas will chose to maintain lower health stocks and will be ill more often than persons in cleaner areas. Proponents of the damage function approach might argue that this is unrealistic since individuals are unlikely to know the precise form of equation (9). All that is necessary, however, for an individual to choose a lower health stock is that he feels less healthy (perceives δ_t to be higher) when pollution increases. Knowing the precise relationshiptbetween δ_t and P_t is irrelevant in choosing H_t^*.

THE VALUE OF A CHANGE IN AIR POLLUTION

We now consider the value to an individual of a small reduction in pollution at time t. Since a change in P_t affects net income only at t the value of a small percentage change in P_t is defined as

$$-\frac{dR}{dP_t} P_t = \left(\frac{d\ lnTL_t}{d\ lnP_t} W_t TL_t + \frac{dI_t}{dP_t} \pi_t P_t\right) e^{-rt}. \qquad (10)$$

The first term on the right-hand-side of (10) is the value of the reduction in sick time caused by a reduction in pollution. This is unambiguously positive. The second term describes the change in investment costs caused by a change in pollution. Reducing pollution increases the optimal health stock which, from (3), increases I*. A reduction in P_t, however, also reduces δ_t which lowers the gross investment necessary to maintain a given health stock. For the functional forms above the net effect of these factors is positive, implying that a reduction in air pollution reduces resources devoted to preventive health care and thus increases willingness to pay,

$$-\frac{dR}{dP_t} P_t = \left(\frac{\alpha\psi}{1+\alpha} W_t TL_t + \frac{\alpha\psi}{1+\alpha} \pi_t \delta_t H^*\right) e^{-rt} = 2\frac{\alpha\psi}{1+\alpha} W_t TL_t e^{-rt}, \qquad (11)$$

If equation (10) is compared with the measure of benefits computed under the damage function approach it is clear that the latter understates willingness to pay. Following Lave and Seskin (1977) the damage function approach would measure the value of the reduction in sick time caused by a reduction in pollution, plus any change in medical costs. Since medical costs are negligible for acute illness, the damage function measure would equal the first term on the right-hand-side of, (10). The second term, which measures the decrease in resources devoted to preventive health care, would be ignored. To indicate the magnitude of this term and to give some idea of the morbidity costs of air pollution we present estimates of (10) based on data from the Michigan Panel Study of Income Dynamics.

ESTIMATION OF WILLINGNESS TO PAY

To compute willingness to pay requires an estimate of $\alpha\psi/(1+\alpha)$, the elasticity of sick time with respect to pollution. Equation (8) suggests that this can be obtained by regressing the log of sick time on the log of pollution and other variables which determine the optimal health stock. Since a large number of persons report zero days of illness each year the appropriate statistical formulation of the equation is a Tobit model,

$$\begin{aligned} lnTL_{it} &= \text{undefined} && \text{if } x'_{it}B + u_{it} \leq 0 \\ lnTL_{it} &= X'_{it}B + U_{it} && \text{if } X'_{it}B + u_{it} > 0 \end{aligned} \qquad (12)$$

where $x_t = (1\ lnPM_t\ lnE_{1t} \ldots lnE_{nt}\ lnP_t\ lnS_t\ lnW_t\ t)$

$$B' = \alpha(1+\alpha)^{-1}(\text{const.}\ 1-\zeta\ -\xi_1\ \ldots\ -\xi_n\ \psi\phi\ -(1-\zeta)\ \tilde{\delta}),$$

and $u_{it} \sim N(0,\sigma^2)$ for all t. Consistent estimates of (12) may be obtained by maximum likelihood.

Table 1 contains estimates of (12) for men between the ages of 18 and 45 from the Michigan Panel Study of Income Dynamics. The dependent variable is days lost from work due to illness, adjusted for differences in weeks worked. Independent variables, apart from the wage, either determine the rate of decay of health capital or affect the productivity of time invested in health.

Two features of the data should be noted. Since the dependent variable cannot be observed for persons too sick to work the estimates in Table 1 are subject to selection bias. This problem is not serious, however, since only 3% of the sample is unable to work for health reasons. Secondly, the data support a threshold model such as (12) since approximately half of the sample reports zero days of illness each year.

Before computing willingness to pay we comment briefly on the performance of the independent variables in Table 1. The first four variables measure factors which affect the rate of decay of health capital--air pollution, pollution at work, parents' income (which may affect δ_o) and race.[57] The first three of these consistently have the expected signs and are significant in six out of eight cases. Race, when significant, implies that being white increases the rate of decay of health capital. The second four variables affect the productivity of time spent investing in health. The presence of a chronic condition has a large negative impact on the productivity of time invested in health and is therefore positively related to sick time. Education, being married and being cautious should increase the prevention received for a given expenditure of resources and are in most cases negatively related to illness.

The chief anomaly in the health equations is the behavior of the wage. A high wage, by increasing the value of healthy time, should increase H^* and reduce TL_t. In Table 1 the wage is either insignificant of positively related to illness. This could be caused by two factors. In the Panel Study the wage is computed by dividing labor income by hours worked. This is not a good measure of the marginal wage unless an individual receives the same wage for each hour worked. Secondly, as Grossman (1972b) has argued, the wage may act as a proxy for deleterious consumption habits, e.g. , eating rich food, which increase the rate of decay of health capital.

We turn now to estimates of willingness to pay. In Table 1 pollution is measured' by the annual geometric mean of sulfur dioxide, which has been linked

TABLE 5.1

HEALTH EQUATIONS FOR MEN 18-45 YEARS OLD[a]

Independent Variable	Interview Year[b] 1970	1974	1976
Constant	"'3.5474	-1.2320	-0.5084
	(1.1253)	(0.9599)	(0.9014)
Ln(SO_2 Mean)	0.2879	0.3168	0.3189
	(0.2140)	(0.2076)	(0.1823)
Works in Manufacturing[c]		0.5001	0.4828
		(0.3659)	(0.3133)
Parents' Income	-0.1832	-0.1310	-0.0150
	(0.0936)	(0.1182)	(0.0953)
Race (1=White)	0.7318	0.3768	-0.2950
	(0.2697)	(0.4052)	(0.3084)
Has a Chronic Health Condition	1.1972	0.6515	0.9347
	(0.4582	(0.2862)	(0.2602'
Yrs. of Schooling	-0.1317	-0.1091	0.0496
	(0.0795	(0.1170)	(0.0508
Marital Status (1=Married)	-0.9678	0.9321	-0.6639
	(0.5098	(0.4550)	0.3828
Risk Aversion Index[d]	-0.3970		
	(0,0881		
Ln(Wage)	0,7492	-0.0899	0.1719
	(0.2873	(0.3553)	0.2813,
σ	2.1460	2.1586	2.1689
	(0.1824)	(0.2656)	(0.1931)
n			

[a]The dependent variable in each equation is the log of [work-loss days/(days worked + work-loss days)]x365. Standard errors appear beneath coefficients.

[b]Each interview year corresponds to the previous calendar year.

[c]Not available in 1970. [d]Not available in 1974, 1976.

Sources: All variables are from the Michigan Panel Study of Income Dynamics except SO_2 which is from the U.S. Environmental Protection Agency.

with acute illness in epidemiological studies. No other pollution variables are included since collinearity between pollutants leads to insignificant coefficients if several variables appear together. SO_2 should therefore be regarded as a pollution index and willingness to pay estimates viewed as indicators of the order of magnitude of willingness to pay. For the interview years 1970, 1974 and 1976 the mean of SO_2 is asymptotically significant at the .10 level or better' (one-tailed test); furthermore its coefficient is approximately 0.3 in each year, despite differences in the specification of the health equation.

Consider now the amount an' individual would pay for an x% reduction in pollution. According to (11) this amount is

$$2(x/loo)\ \frac{d\ln TL_t}{d\ln P_t}\ W_t\ TL_t\ . \qquad (13)$$

In equation (12) the elasticity of sick time with respect to pollution is equal to $\Phi(X'_{it} B/u)$, the probability of being ill, times the coefficient of the log of pollution. Since $\Phi(X'_{it} B/u)$ can be approximated by the fraction of the sample which is ill, $\Phi(X'_{it} B/\sigma) = 0.5$ in each year, implying that the elasticity of sick time with respect to pollution = 0.15. [5] The expected value of TL_t, calculated at the sample mean of X_i, is approximately 40 hours in each interview year. [6]

Equation (13) thus implies that the average person in the 1976 sample, who earned $6.00 per hour, would pay $7.20 annually for a 10% decrease in the mean of SO_2. The damage function approach, by contrast, would put the value of a 10% reduction in pollution at only $3.60. In a city with one million prime-aged men this would understate the value of a 10% reduction in air pollution by $3,600,000 annually. Ignoring adjustments to pollution, therefore, could sizeably understate the value of an improvement in air quality.

REFERENCES

1 For this solution to be valid the resulting value of I_t must lie between 0 and $\bar{I}$, the maximum I permitted at any t. (That $\bar{I}$ exists is guaranteed by the fact that Ω and non-labor income are finite.)

2 It is also true that air pollution affects productivity of time spent exercising; however, not all time invested in health is affected in this way. It therefore seems inappropriate to incorporate pollution in the production function for health.

3 In the paper δ_t is viewed as exogenous, hence the possibility of altering δ_t by moving or changing jobs is ignored.

4 Age, which should also affect the rate of decay of health, was dropped from the equation for lack of significance.

5 Evaluated at the sample mean of X_{it}, $\Phi(X'_{it}B/\sigma) = 0.57$ in 1970; 0.50 in 1974; and 0.53 in 1976.

6 $E(\ln TL_{it}) = X'_{it}B\Phi(X'_{it}B/a) + \sigma\phi(X'_{it}B/a)$. If this expression is evaluated at the sample mean of X_{it}, $E(TL_t)$ is, respectively, 46, 38, and 41 hours in 1970, 1974, and 1976.

BIBLIOGRAPHY

Grossman, M., "On the Concept of Health Capital and the Demand for Health," J. Polit. Econ., 80(1972a), 223-255.

Grossman, M., "The Demand for Health: A Theoretical and Empirical Investigation," National Bureau of Economic Research, New York, 1972b.

Grossman, M., "The Correlation Between Health and Schooling," In Household Production and Consumption, N.E. Terleckyl, cd., Columbia University Press: New York, 1975.

Lave, L.B. "Air Pollution Damage: Some Difficulties in Estimating the Value of Abatement," In Environmental Quality Analysis, A.V. Kneese and B.T. Bower, eds., Johns Hopkins University Press: Baltimore, 1972.

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

U.S. Environmental Protection Agency, Air Quality Data Annual Statistics, Research Triangle Park, Selected Years.

University of Michigan Institute for Social Research, A Panel Study of Income Dynamics, Procedures and Tape Codes, Volumes 2, 4, 5, and 6, 1976.

CHAPTER VI

AIR POLLUTION AND DISEASE: AN EVALUATION OF THE NAS TWINS

INTRODUCTION

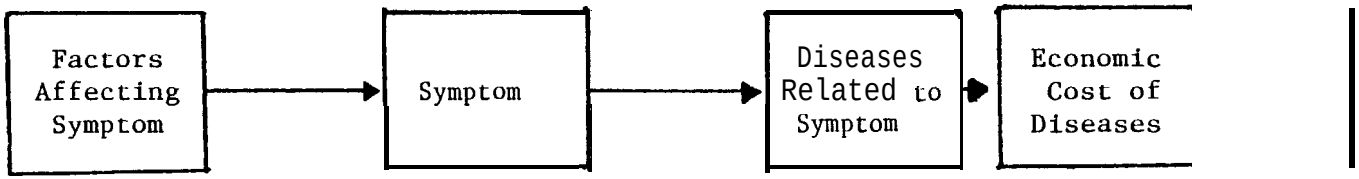
Human disease is caused by a mosaic of events, exposures, psychoses, genetic background, and the environment in which the individual resides. Air pollution is but one of the many factors potentially influencing morbidity and mortality rates of the population. The central question arises as to whether the net effect of air pollution can be assessed and measured such that a scientifically defensible estimate can be made of the change in health resulting from a change in ambient outdoor concentration of air pollutants. In recent years, a number of substantive studies have been undertaken to estimate this net effect. Lave and Seskin (1977) in their monumental work conclude that air pollution, when other factors are taken into account, contributes substantially to increased mortality across cities in the U.S. More recently, Graves and Krumm (1982) have demonstrated a connection (non-linear) between hospital admission rates and concentrations of carbon monoxide and sulfur oxides. Ostro has demonstrated a relationship between work loss days and particulate concentrations. Other studies have connected higher concentrations of air pollutants with indirect measures of lack of health [Gerking (1982).]

In this study we attempt to evaluate the impact of higher ambient concentrations of air pollutants on certain symptoms and reported diseases of a sample of approximately 14,000 twins who served in the Armed Forces during World War II. The simple idea underlying the study is that if there is a relationship between disease and air pollutant exposure, then exposure to higher concentrations of air pollutants, over time, should lead to a higher level of reported symptoms and incidence of certain diseases. Problems arise from many sources in this approach. For example, a symptom such as cough or shortness of breath can be related to the presence of many types of disease, or no disease at all. The presence of a cough, chest pain, and shortness of breath may be caused by asthma, emphysema, chronic bronchitis, or ischemic heart disease, among others. Secondly, the presence of a disease may not be detected because of a lack of one or more symptoms, or not seeking medical treatment. In addition, symptoms may be related to the presence of more than one type of disease. As one illustration, the individual may have both heart arrhythmia and emphysema, and yet exhibit shortness of breath as a single symptom. Finally, symptoms may not be accurately diagnosed and thereby reported on by the individual either because of a lack of basic medical understanding or other reasons. Also, there are substantial difficulties in relating symptoms to the prevalence of diseases, even though symptoms may emerge as a result of higher air pollutant exposures.

Factors other than the presence of air pollutants may have a significant effect on the occurrence of symptoms. Heavy smokers would tend to have a cough and perhaps shortness of breath regardless of air pollution concentrations. Air pollutants would then only exacerbate the presence of the symptom.

These and other qualifications must be kept in mind in evaluating the results reported later. A simple flow diagram (Figure 1) contains most of the hypotheses tested in this study. Examples of the factors proposed to influence the presence of symptoms are given in column 1. The list of symptoms recorded in the National Academy of Sciences twins data set are listed in column 2. A sample of the potential diseases that may be diagnosed from the symptoms are listed in column 3. Finally, in column 4 direct and indirect medical costs are given. In this study, primary efforts were made in relating factors affecting symptoms to symptoms and relating symptoms to the likelihood of a particular disease. As one example, increases in the level of total suspended particulate in the air may cause a greater number of individuals reporting severe chest pain (debilitating for more than one half hour) and shortness of breath when other factors such as cigarette consumption are taken into account. Severe chest pain over a period of time is one of the primary signals of the possibility of coronary heart attack or ischemic heart disease, although the signal may be for something else much less severe. Approximately 2 percent of individuals reporting severe chest pain have a coronary heart attack in the near future. Working through the chain of factors; symptoms, occurence of diseases, and economic cost of diseases, an estimate can be made of the impact of air pollutant exposure on economic costs. From some of the estimates reported later on, a 1 $\mu g/m^3$ increase in total suspended particulate concentration implies a $0.03 Per capita increase in economic costs associated with coronary heart attacks. However, these estimates should be viewed as purely experimental since many of the calculations and assumptions are new and have not been verified or replicated in independent analyses.

In the next section, a brief conceptual economic model is described where symptoms become a part of a household technology in solving medical problems. The following section contains a description of the data set. The next to last section contains the estimated regressions (one set) and final results on economic costs related to air pollutants.



Factors Affecting Symptom	Symptom	Diseases Related to Symptom	Economic Cost of Diseases
Diet	Cough	Chronic Bronchitis	Hospital Expenditures
Age	Shortness of Breath	Emphysema	Loss of Earnings
Asthma	Chest Pain	Bronchiectasis	Loss of Earnings Due to Death
Alcohol Consumption	Severe Chest Pain	Chronic Interstitial Pneumonia	Physicians Services
Income	Coronary Heart Attack	Ischemic Heart Disease	
Cigarette Consumption		Congenital Heart Disease	
Ambient Air Pollution		Cardiomyopathy	
		Cardiac Failure	

Figure 6.1 Major Relationships Examined and Statistically Estimated for the NAS Twins

MODEL DEVELOPMENT

A MODEL OF THE INDIVIDUAL'S HEALTH PROBLEM

It has been said by many people many times before that although they may not be rich, at least they have their health. This not only indicates the importance of one's health in the enjoyment of his life, but further suggests that an individual will normally have more than just a passive interest in the state or quality of his health. Stated in the terminology of the economist, one's health state is a valued good which yields utility to the individual.

There have been a reasonably large number of alternative economic models of health status proposed in the economic literature ranging from lifetime earnings concepts to labor market success. Most of these models concentrate on the effect of health status on the supply or productivity of labor (l). The general conclusion of these studies is that the occurence of diseases may reduce earnings by 20-30 percent through both amount of hours worked and the wage rate received. We have not discovered a study similar to this one which attempts to relate the incidence of disease, through symptoms, to specific causes, such as air pollution. Previous studies by the Wyoming group have focused on sorting out the demand and supply for medical services and how this is effected by air pollution (2). The issue of simultaneity in demand and supply is not addressed in this study.

It is safe to assume 'that an individual would like to have the best quality of health possible, but the procurement of such is not without costs. In particular, the individual may also gain utility from the consumption of goods which will adversely effect his health. For example, he may enjoy smoking cigarettes which has been linked to numerous lung ailments. Thus , the individual must balance his desire for smoking against his desire for good health. The acquisition of better health may also involve the necessary consumption of goods "which in and of themselves yield the individual dis-utility. For example, in order to increase the quality of his health state the individual may have to do some physical exercise when he prefers a more sedentary existence or he may have to eat types and quantities of food which are not to his liking (i.e., a salt-free diet or a simple weight-reducing diet) . Finally, the quest for good health may also involve more direct costs such as medical bills and possibly drugs such as aspirin, vitamins, insulin, or medicines to control blood pressure problems. Hence, one may envision the individual's problem with respect to his health as an economic one where choices must be made and tradeoffs considered between increased health quality and the costs of procuring it. In other words, within limits, an individual's health quality is a variable over which he possesses some control and which he will likely attempt to manage in some optimal fashion. It is the intent of this section to present a model of this problem and the relevant factors which are likely to influence the individual's choice. Particular emphasis will be placed on the role of air quality in this decisionmaking process.

The Utility Function

The utility function of an individual is a relationship between different quantities or bundles of goods and the satisfaction or happiness they provide to the individual in a specified time period. As noted above, the quality of one's health is likely to be a good which yields the individual utility. But numerous others could also be mentioned from French caviar to t-shirts. In this study, however, primary emphasis will be placed on those goods which are likely to either indirectly or directly effect the health of the individual. In particular, the individual's desires with respect to smoking, drinking of alcoholic beverages, nutrition, and the nature of his health state itself. Let the individual's utility function then be expressed as follows:

$$U_t = U_t(Q_t, C_t, B_t, E_t, H_t, X_t) \quad (1)$$

where:

Q_t refers to the air quality levels to which the individual is exposed at time t;
C_t is the quantity of cigarettes consumed at time t;
B_t is the quantity of alcoholic beverages consumed at time t;
E_t is the quantity of exercise (number of minutes) the individual engages in at time t;
H_t is the individual's perceived health status at time t;
X_t is the quantity of a composite good (i.e., all other goods) consumed at time t.

It appears reasonable to assume that the following relationships exist,

$$'_Q' \; U_H, U_X > 0; U_{QQ}, U_{HH}, U_{XX} < 0 . \quad (2)$$

With respect to the other variables, it is possible that either utility or disutility could be generated by the "goods" listed. If the goods are viewed as "goods" by the individual then the following relationships are likely to exist,

$$u_{C'} \; '_{B'} \; U_E > 0; U_{CC}, U_{BB}, U_{EE} < 0 . \quad (3)$$

If they are viewed as "bads" then,

$$'_C' \; '_B' \; U_E < 0; U_{CC}, U_{BB}, U_{EE} > 0 \quad (4)$$

of course, any combination of some of them as "goods" and some as "bads" would also be possible subject to the relationships relevant above.

Several points are relevant to this representation of the utility function. First, the state of one's health appears directly as a source of utility to the individual. It is likely that the health state actually is a joint "input" with the other goods in the "production" of utility but its importance in the utility function should nonetheless be downplayed any more than the role of energy inputs as joint inputs with agent inputs should in

the production of some output. Secondly, although the level of air quality may be viewed as a choice variable of the individual (he can effect it by living in different areas, for example), for the purposes of this investigation it will be taken as given and beyond the control of the individual in order to keep the number of adjustments the individual can make in response to it at a workable level. The inclusion of air quality in the utility function is a proxy for the aesthetic benefits the individual receives from the environment. As air quality deteoriates (i.e., visibility is reduced or the air begins to smell), it is likely that the individual will experience a loss of aesthetic benefits and so, a resulting loss of utility.

Finally, note that the individual may get utility from cigarette consumption which may adversely effect the utility he receives from the quality of his health. Thus, the tradeoff mentioned earlier and the need to more closely specify the nature of the effect on health.

The Respiration Process

In order to understand how various factors influence one's health state it is necessary to gain a rudimentary idea on how the human body works. The normal sequence of chemical changes in human calls depends on oxygen and hence, there exists the need for continuous supply. One of the chief end products of these chemical changes is carbon dioxide and hence, the need for continuous elimination of this waste. In simple single cell animals the intake of oxygen and the release of carbon dioxide occurs at the surface by diffusion. However, as organisms increase in size and complexity, a specialized structure is developed which functions to serve the needs of the various cells. In man this function, known as respiration, is performed by the respiratory system aided by the cardiovascular system.

Oxygen reaches the various cells in the body through three steps: (1) from the environment to the lungs, (2) the lungs to the blood stream, and (3) the blood stream to the cells. The movement of carbon dioxide out of the body is just in the opposite direction. Each of these steps may be discussed separately. The first step, referred to as ventilation, involves inspiration, or the breathing in of outside air and expiration, the breathing out of carbon dioxide. The driving physical force behind this process is Boyle's Gas Law which states that "volume varies inversely with pressure at a constant temperature."

On inspiration the primary muscle of the respiratory system, the diaphragm, pulls downward thus enlarging the cavity containing the lungs. This increase in volume, a la Boyle, causes a reduction in the pressure within this cavity with relative to normal "outside" pressures and so, causes air to rush in and expand the lungs as pressures are equalized. On expiration the diaphragm relaxes and just the opposite occurs forcing air out of the lungs. The substance of the lungs themselves is porous and spongy. Bronchial tubes (hollow air passageways) connect the lungs to the outside environment. Each lung is composed of a large number (billions) of air sacs called alveoli each covered by numerous capillaries. Thus, the ventilation process brings air into these alveoli on inspiration and removes air from them during expira-

tion. The makeup of the air inspired and that expired of course is not the same as that expired in percentage terms as it contains less oxygen (16 percent versus 21 percent) and more carbon dioxide than that inspired.

The second step in the respiration process is called external respiration and involves the passage of oxygen from the alveoli of the lungs to the blood stream (and vice versa, the passage of carbon dioxide from the blood stream into the alveoli). What occurs is the passage of oxygen through the alveoli membrane into the capillaries surrounding it and the opposite passage of carbon dioxide into the alveoli. This transfer occurs due to variances in partial pressures. As noted above, inspired air oxygen makes up a larger percentage of the total volume of air then it, does in the returning blood from the cells and so, has a higher partial pressure. Thus, as blood flows through the capillaries surrounding the alveoli, due to the pressure differentials, oxygen flows from the alveoli into the blood stream. Since the returning blood contains carbon dioxide released from the cells, the partial pressure differential is just opposite and so, carbon dioxide passes from the capillaries into the alveoli where the partial pressure of carbon dioxide is lower. This exchange is influenced by several factors: (1) the area of contact for the exchange, (2) the length of time blood and air are in contact (only about a second or two at any one time--at least once or twice a minute all the blood in the body passes through the capillaries of the lungs), (3) permeability of cells forming the capillary and alveolar membranes, (4) differences in concentrations of gases in alveolar air and the blood, and (5) rate at which chemical reaction takes place between the gases and the blood. Respiratory efficiency is also related to the number of red cells, hemoglobin content of these cells, and the area of the red cell (3).

The final step is internal respiration which involves the passage of oxygen from the blood into the tissue fluid and on into the cells and the reverse passage of carbon dioxide. After the exchange of oxygen and carbon dioxide in the lungs, the newly aerated blood (oxygen-carrying blood) is returned to the heart and then distributed to all parts of the body. As blood moves into the various capillaries, the partial pressure of the oxygen in it is high while that for carbon dioxide is low. Meanwhile, the reverse is true in the tissue fluid and cells since they have "used" previous supplies of oxygen and have created "waste" carbon dioxide. These pressure gradients once again result in the transfer of gases between the blood stream and the cells and thus, complete the respiration process.

The Oxygen Production Function

Given this somewhat brief description of what in reality is a most complex and not fully understood process, the human body, especially the respiratory and cardiovascular systems, may be viewed as a factory which processes an input (air in the environment) into a useful product for the cells of the body (oxygen). There is also the elimination of carbon dioxide, but this may be seen as just another side of the same coin. Considering useable and delivered oxygen to the cells as the output, an economic production function may be envisioned as follows,

$$02 = f(K, A) \quad (5)$$

where:

02 is the amount of oxygen delivered to various cells of the body during a specified time period

A is the total volume of environmental air of fixed quality, Q_t, which is inspired during the specified time period

K is the quality of the individual's "body capital" during the specified time period

In general, it is to be expected that

$$f_A, f_K > 0 \text{ while } f_{AA}, f_{KK} < 0 \quad (6)$$

but a closer examination yields even more information.

It should be clear that the two "inputs" in this production relationship serve different roles. The inspired air is material to be processed by the "body capital" (i.e., the various components of the human body--more on this below) into useable oxygen. Substitution across these two types of inputs may thus only be done up to a certain limit.* For example, if in a sedentary position an individual requires 20 liters of oxygen per hour then clearly at the very least the air inspired during an hour must contain 20 liters of oxygen (actually much more would normally be required since a relatively small percentage of the oxygen inspired is ever taken into the bloodstream). Thus, regardless of the state of the individual's body capital, a minimum of inspired air is required and cannot be substituted for. On the other hand, the body capital must be at some minimum level of efficiency in order to insure the 20 liters of oxygen eventually reaches the cells. So, for any given oxygen requirement during some period there are likely to exist minimum requirements of <u>both</u> inspired air and body capital quality and these requirements will increase with increased oxygen requirements. However, to the extent these minimums are attained some substitution between these inputs are possible. For example, one could achieve a given level of oxygen production in several manners. If the body capital is in a very poor state (but at least the minimum required) this may be offset by a higher flow of inspired air (increasing the rate of respiration). If the body capital is in fairly good shape, clearly less inspired air would be required. These relationships may be represented by the isoquant mapping of this production function shown in Figure 2.

Measured along the vertical axis is increasing body captial quality (measured in terms of some efficiency parameter), while increased quantities of inspired air <u>of a given</u> quality is measured along the horizontal axis. Each isoquant then represents those combinations of body capital quality and volumes of inspired air (again, of a given quality) which would yield a given amount of delivered oxygen to the cells, which as shown, is dependent on the activity level of the individual. Diminishing marginal rates of substitution are assumed. Note that each isoquant approaches both a vertical and horizontal asymptate to reflect the fact that for any level of oxygen produced there exist minimum requirements of both body capital and volumes of inspired air.

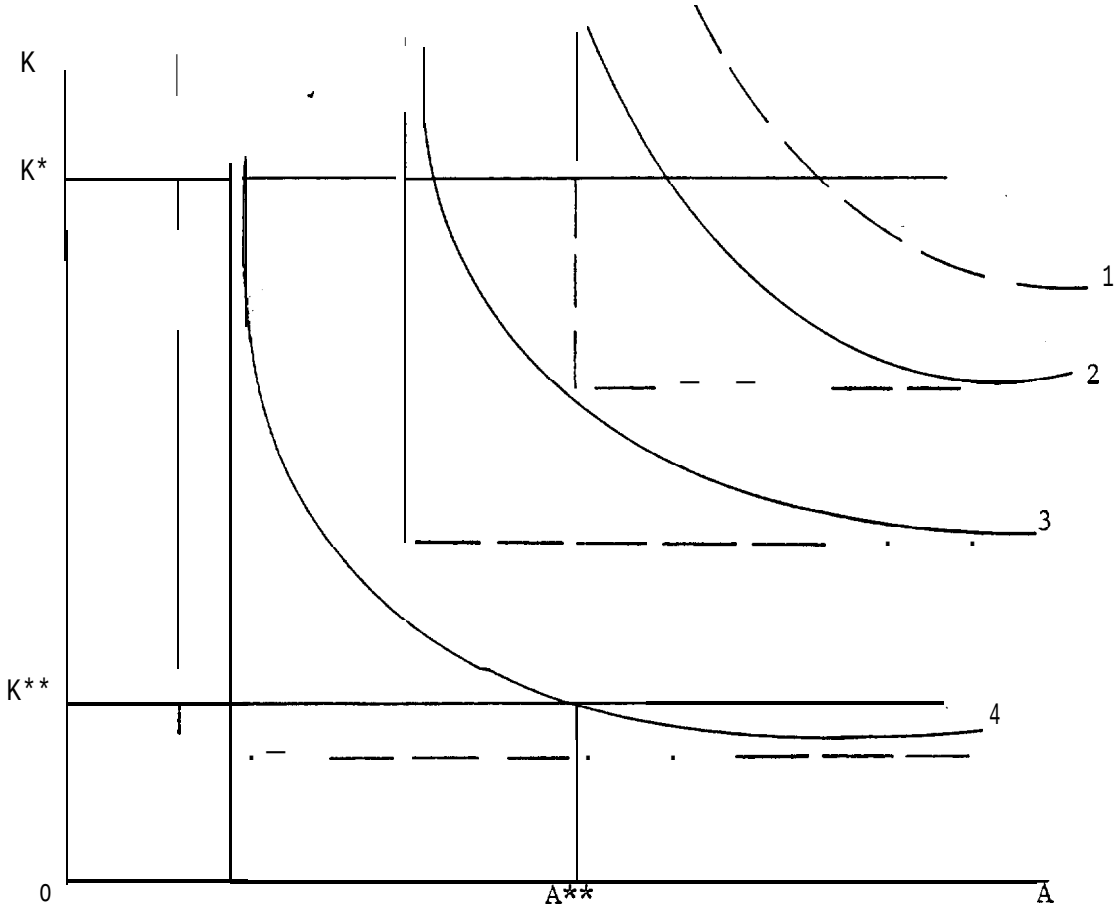


Figure 6.2 Conceptual Tradeoff Between Body Capital and Respiration

[1] O_2 required for heavy physical activity with an inferior air quality

[2] O_2 required for heavy physical activity

[3] O_2 required for light physical activity

[4] O_2 required for sedentary existence

This illustration of the "oxygen production function" of the human body will aid greatly in developing how an individual perceives the state of his health, however, let us digress at this point for a more indepth look at this variable called "body capital".

K or K_t represents the true health status of the individual as given by the quality of his "body capital," that is, the actual physical condition of his heart, lungs, and other components of his respiratory or cardio-vascular systems and the proficiency in which they perform their functions. Though not directly observable by the individual, in general one would expect that

$$K_t = K_t(K_o, \underset{\rightarrow}{Q_t}, \underset{\rightarrow}{C_t}, \underset{\rightarrow}{B_t}, \underset{\rightarrow}{E_t}, \underset{\rightarrow}{X_t}, \underset{\rightarrow}{M^K_t}) \tag{7}$$

where K. represents the individual's initial body capital quality endowment which would be based largely on inherited genes and the subscript $\rightarrow t$ refers to the full "time Profile" of consumption of the respective variable up to time t. This says that not only is the total consumption of some good, say cigarettes, C, important, but also the timing of this consumption. For example, given that an individual's body capital has some natural regenerative capabilities as many feel it does, than one would expect that someone who smoked one pack a day for a year 5 years ago might have a better state of body captial today than someone who smoked a pack a day for the last year. Thus, the quality of one's true health status is probably dependent on cumulative doses, as well as, the timing of those doses. This type of dependence is difficult to model, however, most relevant information may be captured by the following:

$$K_{t+1} - K_t = AK_t = g(K_t, Q_t, C_t, B_t, E_t, X_t, M^K_t) - \delta K_t \tag{8}$$

where K_t would include much of the information concerning past loadings of Q, C, etc. and δ represents a natural decaying factor of the quality of one's body capital with age. Generally it seems reasonable to assume the following,

$$g_Q, g_C, g_B < 0 \text{ and } g_E, g_{M^K} > 0 \tag{9}$$

given the latest medical evidence available (remember, the function g attempts to describe the actual change in one's true health status given a certain level of outside influences and that these true relationships are still not wholly determined by the medical profession). M^K denotes the amount of medical services and/or medicines purchased by the individual to improve the state of his health, i.e., vitamins, medicine to control blood pressure, or simply advice from a doctor. Since X is a "catch-all" including all other goods, it is uncertain how it will over time effect the level of K_t. Finally, included in the behavior of g would be some account for the natural regenerative capability of the body capital. In other words, for levels of Q, C, and B below some threshold level for each, one would expect g to be positive to reflect an improvement in body capital.

The Individual's Perceived Health Status

Given a level of K determined as in (7), let us return to Figure 2. Clearly, if K is at some level such as K* the individual should observe little problem with lack of oxygen. However, if his level of K were more like that of K** then note that light physical activity becomes impossible for him and even a sedentary existence requires more inspired air, A**, then the individual with K* quality (A*). This second individual will thus be getting a symptom (i.e., shortness of breath or chest pain if his heart must do extra duty to process more air) that something is wrong.

Another manner in which a symptom, a physical response of the body, might occur involves the level of air quality. However, suppose the air quality was worse. For a lower level of air quality it is likely that the isoquants of Figure 2 would shift in a northeasterly direction. That is, to produce a given amount of delivered oxygen would require both more inspired air (since the useable portion of this air would be less) and a higher quality of body capital since more of the material input would have to be processed. This suggests that an individual with a given level of K may experience no symptoms in a "good" air quality situation, but as air quality deteoriated symptoms would arise as the minimum requirements of inspired air rose.

Given the above, a symptom, an observable phenomenon to the individual, has basically two sources--a deteoriation of body capital or a deteoriation of air quality. With respect to air quality then it is possible to distinguish between its chronic effects (its effects on the quality of body capital) and its acute effects (its effects on changes in the useable nature of the material input--inspired air). So, the advent of a symptom may be the result of a true deteoriation of health status or simply the result of deteoriating environmental quality (wherein health status is actually not in jeopardy). Take coughing for example. This symptom could occur because the quality of body capital has been reduced to low levels and so even with good quality air the individual coughs (for example, the individual could be a long-time smoker and this has led to emphysema wherein many of the alveoli of the lungs have been rendered all but unusable). On the other hand, coughing could occur because of a high concentration of some pollutant in the air one breathes (that is, the individual's health status may be okay, but the material input of the oxygen producing process is in some manner inadequate or unusable). Of course, the coughing could also be a result of both inferior quality body capital and inferior air quality. In any case, it is likely that

$$St = St\ (K_t,\ Q_t,\ M_t^s) \qquad (10)$$

or that the occurence of some symptom is dependent on the true state of the individual's health, air quality, and possibly on medicines used to alleviate the advent of a symptom (i.e., one could use cough drops to reduce coughing, eye drops to reduce eye irritation, or aspirin to relieve a headache). Given this it is likely that

$$'K' \quad 'Q_t+' \quad {}_{iM}^{S}s > 0_{\text{"}} \tag{11}$$

These symptoms are the only observable manner in which the individual may get a perception of his true health state. If there are no symptoms to the contrary an individual is likely to assume he is okay while if some are prevalent he is likely to assume that something is not right. Another way in which he may evaluate his health status is to procure medical information. For example, although a person with high blood-pressure rarely has noticeable symptoms, a blood pressure test could reveal the problem and thus, give the individual a clearer picture of his health status. Also, going back to the example of coughing above, a medical check-up could tell the individual if in fact the coughing was due to something like emphysema or instead just by "something in the air" meaning his health state was okay. This suggests that

$$H_t = H_t\,(St,\ M_t^I) \tag{12}$$

or that the individual's perceived health status depends on the symptoms he observes and any additional medical information he has purchased concerning how to evaluate these symptoms or discovering health problems without current or may assume he is okay and that there is merely "something in the air" depending on his opinion and that of any medical person. In either case, his behavior will be based on his perception of his health status whether or not this perception is right or wrong. That is, an individual behaves according to the perceived state of his health and not the actual or true state. Mathematically, the individual's health problem may be stated in continuous terms as follows:

$$\max_o \int^T U(Q,\ C,\ B,\ E,\ H,\ X)e^{-rt}dt \tag{13}$$

subject to:

$$\dot{K} = g(k,\ q,\ C,\ B,\ E,\ X,\ M^K) - \delta K$$

$$s = S(K,\ Q, M^S)$$

$$H = H(S,\ M^I)$$

$$\bar{Y} \geq P_X X + P_C C + P_B B + PEE + P_M(M^K + M^S + M^1)\ \forall t$$

$$K(0) = K_0$$

where $\bar{Y}$ is the individual's income constraint and P are the various prices of the respective marketed goods. This is an optimal control problem wherein the individual's health state and his consumption of other commodities act as control variables and his true health state, K, is the state variable with its equation of motion. In other words, the individual's problem involves manipulating C, B, E, H, and X subject to a budget constraint in order to maximize his utility. A solution to this model will depend on what assumptions

are made (is $U_{Q} \gtrless 0$?) but the important tradeoffs will be adequately represented. Further note that the model allows for all three manners in which a change in air quality might effect the utility of an individual: (1) directly through aesthetic effects, (2) indirectly through changes in his body capital which will effect his health status and finally, (3) indirectly through changes in the symptoms he may observe which again effect his perception of his health status.

An important step towards the solution of this model involves the link between air quality, cigarettes, etc. and the advent of symptoms or an estimation of the symptom function, S_t. This is a primary objective of the remainder of this study.

Unfortunately, a thorough search of the medical literature has revealed practically no applicable equations to estimate even a "proxy" for health status or "body capital," or for the oxygen production function. In consequence, we have had to abandon this modelling approach and apply a more simple model structure.

Outline of the Model Applied

It has been proposed in many economic studies of health effects that individuals derive disutility from perceived and/or actual occurrences of disease. However, most individuals cannot correctly diagnose their own diseases except for a small set of common ailments. The individual commonly perceives one or more symptoms of the potential occurence of a disease. The individual may then select three alternatives, to seek medical services for diagnosis and cure; to use self-prescribed medication or other forms of self-help, or to do nothing. Typically, the individual will make these choices based on the severity of symptoms and the cost of medical services. If the symptoms are common types, i.e., the sudden appearance of a slight chest pain, the individual is likely to do nothing. Also, if the cost of medical services is extremely low or negative, the individual is likely to seek medical attention for the appearance of any symptom. The important point is that individuals work with symptoms and not the actual disease itself, whether it is the afflicted party or the physician making the diagnosis. Thus, we postulate a simple welfare relationship where S denotes a vector of symptoms and I a vector of other goods and services the individual purchases. Then the individual's utility can be represented as:

$$u = U(x, s) \tag{14}$$

where, for illustrative purposes, the function u(*) is assumed to be continuous in I and S and twice differentiable. The individual is assumed to be constrained by a budget constraint on purchases of medical services to alleviate symptoms or cure diseases and purchases of other goods and services:

$$P_X X + P_M M \leq Y \tag{15}$$

where M is the quantity of medical services, Y is income, and P denotes the unit price of the service X either as a scalar or vector. finally, to

complete this simple model, we denote a relationship between the incidence and severity of symptoms and required medical services. For simplification, it is assumed there are a fixed set of medical services to alleviate symptoms or treat various diseases, provided the individual seeks treatment and that this relationship can be expressed as:

$$M = h(S) \tag{16}$$

Next it is presumed the individual maximizes utility subject to the budget constraint and medical technologies. The first order conditions become:

$$\left.\begin{aligned} {}'_X - \lambda P_X &\leq 0 \\ {}'_S + \delta h_S &\leq 0 \\ -\lambda P_M - \delta &\leq 0 \end{aligned}\right\} \quad {}'_S + \lambda h_S \leq 0 \tag{17}$$

with $\lambda \geq 0$, $\delta \geq 0$, $u_X \geq 0$, and $u_S \leq 0$. These conditions simply indicate that the maximizing individual will purchase goods and services up to the point where marginal utility for goods is equated with the utility adjusted price of the goods. The individual will purchase a reduction in symptoms (improvement in health) up to the point where marginal disutility associated with symptoms is equal to utility adjusted productivity of purchases of medical services. Note that this follows regardless of whether there is a correct doagnosis of symptoms. What is important to the individual is whether the symptoms are alleviated and a return to good health status is perceived. A derived demand relationship for M can be developed from the presence of symptoms as follows:

$$M = f(P_X, PM, s) \tag{18}$$

where $f(\cdot)$ evolves from the first order conditions in (17). Following Mäler (1974), compensating and equivalent variation measures of consumer surplus can be constructed for S where the individual cannot control the appearance of symptoms except through changes in lifestyle or preventative actions which will not be considered here. While conceptually willingness to pay to avoid symptoms or associated medical expenses can be derived, no attempt is made in this study to estimate equation (18). The reason for not doing so is that no adequate data exist for the NAS twins to estimate M or PM. As an alternative, average U.S. medical expenditures for each type of illness were used to estimate a minimum willingness to pay to avoid symptoms . The underlying assumption is that individuals, at minimum, would be willing to pay to avoid symptoms what they typically do pay to alleviate them. In this sense, a minimum estimate is calculated.

THE DATA SET

NATIONAL ACADEMY OF SCIENCES TWIN REGISTRY*

The data which 'this research analyzes to discover the net effects of air pollutants was obtained from the NAS-NRC Twin Registry (4). This twin panel consists of 7,960 white male twin pairs, of which 6,741 twin pairs or less are examined in this study. Table 1 summarizes the age distribution of the NAS Twin panel in 1967 when the panel was asked to complete the epidemiological questionnaire (Q2) which provides the relevant health data. The twins ranged from 41 to 51 years of age at the time the Q2 information was collected. The average age was 45.

The sample itself is the result of a detailed procedure by which the National Research Council identified white male twins born during the period 1917 to 1927 in the continental United States. Additional screening was done on this set of twins to determine the twin pairs for which both members served in the armed forces (5). The process resulted in the 7,960 twin pairs currently comprising the Twin Registry.

An initial questionnaire (Q1) was used to obtain each individual's medical history since separation from military service and to identify the brothers zygosity (6). Figure 3 presents the question used on Q1 to obtain each individual's medical history since military separation. This information provides the basis for a diagnostic index which is maintained for the NAS-NRC Twin Registry. This Q1 information has been updated and purged from the diagnostic index as more complete information in medical history was collected based on Veterans Administration (VA) claims records, VA hospital records, and death certificates. In fact, the present diagnostic index is largely based on such VA information sources rather than the self-reported information from Q1.

The reader might find it tempting to consider using information in the diagnostic index to quantitatively define health status in the sort of statistical exercise which is summarized below. However, the diagnostic index represents an amalgam of different data sources each of which would be expected to contribute its own unique biases to such an analysis. For example, the self-reported Q1 information is purged when VA information is available. Therefore, the entire set of VA criteria determines the set of Q1 information that remains. Fundamentally, the VA criteria relate to military causes of medical problems as well as a certain socio-economic status. Actual information in the diagnostic index, because it is collected from different sources, may be inconsistent and therefore potential introduction of biases is difficult if not impossible to sort out.

TABLE 6.1 AGE DISTRIBUTION OF NATIONAL ACADEMY OF SCIENCES TWIN SAMPLE - 1967

Age	Absolute Frequency	Relative Frequency	Cumulative Frequency
41	1622	12.0%	12.0%
42	1646	12.2	24.2
43	1470	10.9	35.1
44	1536	11.4	46.5
45	1419	10.5	57.1
46	1265	9.4	66.4
47	1282	9.5	76.0
48	1180	8.8	84.7
49	786	5.8	90.5
50	744	5.5	96.1
51	532	3.9	100.0
TOTAL	13,482	100.0	

List **any** illness, impairment, disability, hospitalization, and operation you have had since separation from military service, stating the year when it first occurred.

Illness, impairment or operation	Year it began	**Name of Hospital**	City **and** State

Figure 6.3 **NAS Twins (Ql) Self-Reported Medical** History **Questionnaire**

And Now Some **Rather Specific Questions About Where You Have Lived Since the Second World War**

50. For consecutive periods, fill in length of period, city or community, as well as state.
Check also at the right of Table in what type of area you were living and working, respectively.

PERIOD OF TIME	CITY OR TOWN	STATE	LIVING IN			WORKING IN		
			Downtown Area	Suburban Area	Rural Area	Downtown Area	Suburban Area	Rural Area
1945 -			☐	☐	☐	☐	☐	☐
			☐	☐	☐	☐	☐	☐
			☐	☐	☐	☐	☐	☐
			☐	☐	☐	☐	☐	☐
			☐	☐	☐	☐	☐	☐
			☐	☐	☐	☐	☐	☐
			☐	☐	☐	☐	☐	☐
			☐	☐	☐	☐	☐	☐

Figure 6.4 **NAS Twins (Q2) Residence and Work History**

The epidemiological information obtained in 1967 from Q2 is the basis for the quantitative measures of health status that are utilized in the statistical analysis which is summarized here. The Q2 health status information is separated into information on respiratory and cardiovascular health problems.

Information on respiratory health status is provided by answers to two questions: do you get short of breath walking with other people at an ordinary pace on the level? Do you regularly or for extended periods of time have a cough? Clearly the binary answers to these questions are either yes or no.

With respect to cardiovascular health status a series of three binary questions provide relevant information. Have you ever had any pain or discomfort in your chest? Have you ever had a severe pain across the front of your chest lasting for a half hour or more? Have you ever had a heart attack?

The statistical analysis summarized later uses the answers to these five questions as binary dependent variables in a regression analysis. Q2 also provided information on a number of potentially relevant explanatory variables. The individual is asked by Q2 to report if he has ever had asthma, his height and weight, whether he has to diet to keep his weight down, the number of cigarettes and cigars smoked per day, as well as the individual's alcohol consumption. In addition, Q2 collects relatively detailed information on dietary habits.

A particularly interesting set of information collected by Q2 is a detailed residence and work history by location. Figure 4 presents the question used to gather this information. This type of information may be particularly useful to a statistical analysis examining the association between air pollution and human health not only because it identifies past residences by city and state, but also because it identifies if the residence and work location were in a "downtown", "suburban", or "rural" area.

Finally, a third questionnaire (Q3) collected economic information such as household income. Unfortunately, Q3 was completed by the panel in 1973 rather than 1967 when the Q2 health information was obtained. Yet Q3 provides the only economic information and 1973 household income is used as a proxy for the same 1967 variable in the statistical analysis. The actual income question was: "How much was your family income from all sources (during 1973)?"

Q3 also provided information on an individual's access to medical care. Q3 asks a detailed set of questions relating to whether the individual does or does not have an annual medical check-up. If SO, additional information is gathered on the source of payment of check-up: government clinic, union clinic, company clinic, or medical insurance.

Air Quality Data

The United States Environmental Protection Agency maintains air quality data information for approximately 12,000 sites. Presently only about 4,000 sites are operational (7). Prior to 1972, air quality measurements were net undertaken on a large scale, and were often subject to considerable measurement errors. The EPA data are published annually in Air Quality Data - Annual Statistics. The air quality data used in the statistical analysis presented below is from this 1977 annual publication.

Air quality data was matched to individual data from the NAS Twins Registry by three digit zip code. The most disaggregated measure of air quality was found to be based on three digit zip codes. Five digit zip codes were not a useful basis for air quality data collection because the number of correspondences between air quality monitoring sites and five digit zip codes was minimal.

The data actually collected by three digit zip code included: maximum 24 hour measurement for total suspended particulate and sulfur dioxide; and type of monitoring station.

Frequently it was necessary to choose between a number of monitoring sites as representing air quality measurements for a given three digit zip code. The criteria by which such decisions were made were: (1) discard all sites for which measurements were discontinued before the end of the year 1977, (2) discard all sites which were not identified by type of monitoring station, (3) choose that site which measures the largest number of pollutants, (4) if two or more sites measure the same number of pollutants, choose the site which has operated the longest, (5) if a choice cannot be made, choose the site with the largest number of measurements for total suspended particulate and (6) if a choice still cannot be made, choose randomly. Note that these criteria were to be applied in sequence from first to sixth.

The importance of the monitoring station type is with respect to matching the air quality data to individual twin registry data. It was pointed out in discussing Figure 4 that residence and work history information was obtained by Q2 with reference to urban, suburban, or rural locations. Similarly, air quality monitoring stations are identified as being located in "center city", "suburban", or "rural". Therefore, air quality data collection was based both on three digit zip codes and on the urban, suburban, rural classification. For each three digit zip code, the goal was to find an urban, suburban, and rural measurement. Unfortunately, this was not always possible. Finally, the actual combination of health data with air quality data has been accomplished by matching the most recent individual residence urban-suburban-rural location by three digit zip code and with the appropriate urban-suburban-rural three digit zip code air quality data.

Unfortunately, the various data sets apply to different points in time in that symptoms increase, and air pollution concentrations are measured in 1967, 1973, and 1977, respectively. In addition, there are difficulties in relating long term air pollutant exposures to individuals at the last location they have resided at. More than one half of the twins have resided since 1945 in two or more locations, and it is unlikely that ambient concentrations in the different locations would be comparable. A second qualification is that cumulative estimates of cigarette or alcohol consumption have not been calculated. In consequence, current non-smokers may have symptoms but have no current cigarette consumption.

STATISTICAL RESULTS

In this section, a reasonably meaningful sub-set of the statistical results are presented along with a partial interpretation of their meaning. The data set after calling out observations with incomplete data or unusable responses to questions ended up being between 7,892 and 7,908 in number. This represents slightly more than 50 percent of the original NAS twins data set. Most of the deletions were due to the inability to obtain matching zip codes between the living location of the twins and an air monitoring station. The bias resulting from this omission is not known. However, it can be anticipated that most of these omissions are of twins residing in suburban or rural locations without monitoring stations, in which case there are fewer observations on those exposed to lower ambient air pollutant concentrations. The effect is to give less dispersion to exposures and thereby insert an indeterminate bias on the estimated coefficients and make their significance less than would be the case.

Given the lack of dispersion in age, socio-economic class, and race we should also anticipate a bias downward in estimated effects of air pollution exposures as contrasted to the U.S. total population. The relative uniformity of the NAS twins sample reduces problems of bias associated with comparing non-homogeneous groups and unknown group differences but increases the liklihood that nothing will be detected connecting air pollution to symptoms of disease when in fact there is a connection.

With these qualifications in mind, we now turn to the actual results. In Table 3 are recorded the means and standard deviations of the variables examined. In Table 4, a raw correlation matrix of results is presented for all of the variables. There is very little correlation between most of the variables with two notable exceptions. There is substantial correlation between the various measures of nutrients and minerals consumed. For example, the raw correlation coefficient between sugar and unsaturated fatty acids consumption is .75. While the correlation coefficient between calcium and vitamin A consumption is .84. Relatively high correlations were also observed among symptoms, which might be anticipated in that severe chest pain is a form of chest pain ($r = 0.32$) and cough and shortness of breath may occur simultaneously, ($r = 0.19$). For the remainder of variables, there is little or no raw correlation which would be expected of a relatively homogeneous data set of 8,000 observations.

Evaluation of Statistical Results from Regressions

After some preliminary experiments with the NAS twins data set, several conclusions emerged. First, there was no effect of the Twins on the estimated relationship between prevalence of a symptom and exposure to air pollution. Thus, there appears to be no discernible “genetic” effect at least in the sample analyzed. Second, a variable reflecting zygosity of twins was never even marginally significant. However, much more detailed statistical comparisons would need to be made in order to rule out the possible

TABLE 6.2 DEFINITION OF VARIABLES*

TWNO	Twin Number	Number of twin
CHPN	Chest Pain	Whether the individual experienced chest pain in 1967, (yes or no)
SHBR	Shortness of Breath	As measured by self-reported statement as to whether it was encountered when walking with friends, in 1967
COGH	Cough	Whether or not the individual regularly or for extended periods of time had a cough before or during 1967
SVCP	Severe Chest Pain	Whether the individual experienced severe chest pain lasting one half hour or more in 1967, (yes or no)
CORN	Coronary	Whether or not the individual had suffered a heart attack before or during 1967
ASTM	Asthma	Whether the individual had asthma before or during 1967, (yes or no)
RHMF	Rheumatic Fever	Whether or not the individual had rheumatic fever or rheumatic heart disease during or before 1967
DIET	Diet	Whether the individual undertook a diet for excess weight before or during 1967, (yes or no)
SMKN	Smoking	Cigarette consumption (packs per day) where conversions are used for cigars and pipe smokers before or during 1967
DRNK	Drunk	How often did the individual drink at least one pint of liquor or two bottles of wine or four quarts of beer at one occasion in 1967
INTX	Intoxica-tion	Whether or not the individual becomes intoxicated daily, in 1967
CTRM	Cigarette Tar	Tar from cigarettes in milligrams per year, in 1967
LIQR	Liquor	Alcohol consumption, beer, wine, and spirits converted to ethanol equivalents in oz. per year, in 1967
HGHT	Height	Height in inches, in 1967
WGHT	Weight	Weight, in 1967

TABLE 6.2 (continued)

WT25	Weight at Age 25	Weight at age 25
BRTH	Birth	Year of birth
EARN	Earnings	Family earnings in 1973 (dollars)
TSPM	Maximum Total Suspended Particulates	Maximum 24 hour concentration in 1977, in $\mu g/m^3$
SO_2M	Maximum Sulfur Dioxide	Maximum 24 hour concentration in 1977, in $\mu g/m^3$
ZYGT	Zygosity	Classified as either monozygotic for identical twins and dizygotic for fraternal twins
STFT	Saturated Fatty Acids	Grams per year, in 1967
SUGR	Sugar	Grams per year, in 1967
FIBR	Fiber	Grams per year, in 1967
USFT	Unsatruated fatty acids	Grams per year, in 1967
NTRS	Nitrosamines	µg per year, in 1967
IRON	Iron	mg per year, in 1967
CALC	Calcium	mg per year, in 1967
THMN	Thiamin	mg per year, in 1967
NIAC	Niacin	mg per year, in 1967
VITA	Vitamin A	lu per year, in 1967
FATS	Fats	Grams per year, in 1967
PROT	Protein	Grams per year, in 1967
RIBF	Riboflavin	mg per year, in 1967

*Tables documenting conversions for food intake variables are reported in Appendix 1.

TABLE 6.3 MEANS AND STANDARD DEVIATIONS OF VARIABLES

Variable	Mean	Standard Deviation
CHPN	.24861	.43920
SHBR	.07145	.41695
COGH	.11292	.33212
SVCP	.04906	.21602
CORN	.11596	.85695
ASTM	.12355	.60376
RHMF	.03541	.18482
DIET	.22129	.41514
SMKN	.60255	.51997
DRNK	.85559	.38287
INTX	3.0567	14.208
CTRM	134.87	217.42
LIQR	425.37	643.76
HGHT	69.783	2.5466
WGHT	172.14	22.056
WT25	158.81	20.808
BRTH	22.956	2.9229
EARN	6.1792	11.687
TSPM	129.54	144.29
SO_2M	49.594	88.189
ZYGT	1.5622	.55714
STFT	7.5156	2.8017
SUGR	51.575	13.022
FIBR	.82409	.28507
US FT	8.6647	3.9050
NTRS	.07108	.06183
IRON	2.3021	.71629
CALC	.25948	.01400
THMN	.34392	.14039
NIAC	2.9235	.86349
VITA	.46757	.19714
FATS	17.436	6.7949
PROT	18.606	4.9021
RIBF	.57155	.08897

TABLE 6.4 CORRELATION MATRIX

	RIBF	PROT	FATS	VITA	NIAC	THIN	CALC	IRON	NTRS	USFT	FIBR	SUGR	STFT	ZYGT	SO_2M	TSPM	EARN	BRTW	WT25	WGHT	HGHT	LIQR	CTRM	INTX	DRNK	SMKN	DIET	RHMF	ASTH	CORN	SVCP	COGH	SHBR	CH	PN	TWNO
1 WNO	-0.00	-0.00	-0.00	-0.00	-0. 00	0.00	-0.00	-0.00	0.00	-0.00	0.01	-0. 00	0 00	0.01	0.00	0.01	-0.02	-0.03	-0.01	0.01	-0.00	-0. 01	-0.04	-0.02	-0.01	-0.03	0.00	0.(31	-0.01	0.01	0.00	-0.01	-0.01	0.00	I.	(MI	
CHPN	-1.01	-0.01	-0. 01	-0.01	-0.01	-0.02	0.00	-0. 01	-0.02	-0. 00	-0.02	0.01	-0. 00	-0. 00	0.02	-0.00	-0.03	0.00	-0. 05	-0.04	-0.02	0.05	0.06	0.04	0.02	0.04	0.04	0.01	0.07	0.20	0.32	0.22	0.19	1.00		
SHBR	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.01	-0.02	0.01	-0.02	-0.01	0.01	0.01	-0.02	0.01	-0.04	-0.01	-0.01	0.01	-0.01	0.0S	0.08	0.07	-0.02	0.05	0.02	0.00	0.06	0.29	0.18	0.19	1.00			
COGH	-0.01	-0.01	-0. 01	-0. 01	0.01	-0.00	-0. 01	-0. 01	-0.01	-0. 00	'0.06	0.01	-0.01	0.02	-0. ON	0.02	-0.04	-0.01	-0.03	-0.04	-0.03	0.15	0.27	0.01	0.07	0.17	-0.03	-0.01	-0.06	0.04	0.12	1.00				
SVCP	-0.01	-0.00	-0.01	-0.01	-0.01	-0.02	-0.01	-0. 01	-0.02	-0. 00	-0. 01	-0.01	-0.00	0.01	0.02	0.01	-0.02	-0.01	-0.03	-0.03	-0.02	0.00	0.05	0,02	-0. W	0.02	0.05	-0.01	0. 00	0.15	1.00					
CORN	-0. 05	-0.04	-0.04	-0.04	-0.04	-0.05	-0.04	-0. 05	-0.02	-0. 04	-0.02	-0. 01	-0.03	0.01	-0.01	0.03	-0. 03	-0. 05	0.01	-0. 02	-0.02	0.01	0.03	0.02	-0.02	0.00	0.08	0.01	0.01	1.00						
AS TH	0.01	0.02	0.02	0.01	0.02	0.02	0.01	0.02	0.03	0.01	43.00	-0.00	0.01	-0.00	0.01	0.00	0.02	-0.01	-0.01	-0.00	0.03	0.01	0.00	0.02	-0. 00	-0.01	0.01	0.00	1.00							
RHMF	-0.02	-0.02	-n. 01	-0.02	-0.01	-0.03	-0.01	-0.03	-0.02	-0.01	-0.01	0.00	-0.01	0.01	-0.01	-0.01	0.02	0.02	0.00	0.00	0.01	0.02	-0.01	-0.012	-0.00	0.01	0.02	1.00								
DIET	-0.05	-0.04	-0.05	-0.03	-0.09	-0.07	-0.07	-0.06	-0. 00	-0.05	-0.06	-0. 10	-0.04	0.00	0.03	0.04	0.09	'0.02	0.21	0.29	-0.00	0.02	-0. 08	0.00	0.01	-0.07	1.00									
SMKN	0.01	0.02	-0.00	-0.01	0.05	0.03	-0. 00	0.02	-0. 00	-0. 00	0.01	-0. 00	0.00	0.00	-0. 00	0.00	-0.04	0.03	-0.01	-0.07	-0.00	0.17	0.49	0.08	0.21	1.00										
DRNK	-0.10	-0.08	-0. 10	-0.09	-0. 06	-0.09	-0.09	-0.09	-0. 04	-0.10	-0.01	-0.07	-0.09	-0.02	0.03	0.01	0.05	0.02	0.00	-0.00	0.02	0.26	0.13	0.09	1.00											
INTX	-0 02	0.01	-0.01	-n. 02	-0. 00	-0.01	-0.02	-0.01	-0.01	-0.02	-0.06	-0. 04	-0.01	0.01	-0. 00	-0.02	-0.01	0.03	-0.01	-0. 00	-0. 01	0.32	0.14	1.00												
CTRM	-0.01	0.00	-0.01	-0.02	0.04	0.02	-0.03	0.00	0.00	-0.02	-0.04	-0.03	-0.01	0.01	-0.00	0.01	-0.07	0.05	-0.02	-0.07	-0.00	0.21	1.00													
LIQR	-0.11	-0. 07	-0.11	-0.10	-0.03	-0.05	-0.13	-0.07	-0.02	-0.12	'0.05	00.14	-0.10	0.00	0.02	-0.02	0.02	0.01	0.01	0.05	0.06	1.00														
HGHT	0.02	0.03	0.02	0.01	0.03	0.04	-0.01	0.03	0.01	1.01	0.01	-0.01	0.01	0.06	-0.(51	0.01	0.05	0.05	0.51	0.53	1.00															
WGHT	0.01	0.01	O.(I1	N.(11	-0.01	0.01	-0.00	0.01	0.03	0.01	-13.02	-0.01	0.01	0.03	-0.01	0.01	0.04	0.01	0.73	1.00																
WT25	-0.(32	-0.01	-0. 00	0.00	-0. 03	-0.01	-0.02	-0. 01	0.02	-0. 02	'0.0^2	-0.05	-0.01	0.05	-0. 02	0.01	0.04	0.05	1.00																		
BRTH	-0.01	-0.01	-0.01	-0.02	0.00	0.01	-0.02	-0.01	0.07	-0.01	0.02	0.01	-0.01	-0.03	0.01	0.01	-0.00	1.00																		
EARN	-0.01	0.00	-0. 00	-0.01	-0. 01	-0.01	-0. 02	-0.00	-0.00	-0.01	0.01	-0.03	0.00	-0.06	-0. 00	0.02	1.00																			
TSPM	-0.01	-0.00	-0.00	-0.02	0.00	-0.00	-0.02	-0.01	0.00	-0.01	-0. 00	-0.02	-0.00	-0.02	0.22	1.00																				
SO_2M	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.01	-0.02	-0.01	-0.02	0.01	-0. 00	-0.02	-0.00	1.00																					
ZYGT	0.01	0.01	0.02	0.01	0.01	0.01	0.02	0.01	-0.01	0.02	-0.01	0.01	0.01	1.00																						
SUFT	0.90	0.90	0.93	0.64	0.76	0.59	0.80	0.86	0.21	0.9?	0.20	0.>1	1.00																							
SUGR	0.68	0.50	0.62	0.40	0.43	0.43	0.34	0.49	0.1S	0.75	0.40	1.00																								
FIBR	0.28	0.29	0.21	0.19	0.19	0.28	0.31	0.36	0.12	0.19	1.00																									
USFT	0.87	0.86	0.95	0.66	0.71	0.58	0.87	0.83	0.20	1.00																										
NTRS	0.33	0.39	0.42	0.14	0.61	0.?6	0.21	0.35	1.00																											
IRON	0.92	0.98	0.89	0.79	0.81	0.79	0.80	1.00																												
CALC	0..97	0.79	0.84	0. 84	0.56	0.51	1.00																													
THIN	0.)0	0.79	0.74	0.46	0.81	1.00																														
NIAC	0.70	0.239	0.81	0.40	1.00																															
VITA	0.85	0.72	0.67	1.00																																
FATS	0.86	0.93	1.00																																	
PROT	0.91	1.00																																		
RIBF	1.00																																			

presence of a genetic effect. Since no "genetic" effect was observed, the researchers decided to "pool" the usable twin observations for further statistical analysis. Third, ordinary least squares and probit statistical computations were made on the same data and no difference was observed in estimated coefficients or their standard errors. In consequence, statistical estimated concentrated almost exclusively on application of the ordinary least squares technique. Finally, it was observed that using a randomly drawn sample of twins to estimate the coefficients (of about 5 percent of the population) yielded coefficients in another. This suggests that for prediction purposes and accuracy, the entire population should be used for estimation purposes.

With approximately 8,000 unique and usable observations, it can be expected that R^2's will be relatively low and that was what was observed uniformly throughout the results.

In Table 5, are recorded the four variants of the regression equation for chest pain. The second equation is the same as the first except intake of sugar is added. For the third variant saturated fats is added, and in the fourth variant, vitamins, proteins, and minerals are added. Across the four variants, none of the independent variables' coefficients or "t" statistics changes very much. And the R^2's are uniformly low. The statistically significant variables are smoking, liquor consumption, but not heavy drinking, earnings, sugar intake, and to a lesser extent, maximum 24 hour concentrations of $S0_2$. As would be expected, smoking contributes to increased levels of chest pain (8). The most common mechanism would be smoke ingestion requiring more inspiration/expiration for the same level of oxygen and thereby greater requirements on the heart for pumping. Greater daily consumption of alcohol stresses the cardiovascular system so it is expected that this would have a positive effect on the incidence of chest pain (9). Birthdate or age has no impact, but this is to be expected given the sample age only ranges from 41 to 51 years. Earnings have a significant negative effect on chest pain. In this equation, earnings probably reflect education and knowledge of diseases and the demand for medical services plus other socio-economic effects. Thus, no economic interpretation (in demand and supply terms) can be made of the earnings coefficient. Finally, while the TSP coefficient is insignificant, the $S0_2$ coefficient is significant at the 95 percent confidence level, and remains stable in magnitude across the four variants of the regression. The coefficient indicates a one ten thousandth increase in the probability of chest pain given a 1 $\mu g/m^3$ increase in maximum average 24 hour concentrations of $S0_2$.

Table 6 contains the estimates for four variants depending on dietary specifications for the symptom, severe chest pain. Again, as with chest pain, smoking and whether the individual had dieted were statistically significant at the 97.5 percent level. Neither air pollution variables were significant across the four variants. Earnings again were negatively significant at the 95.5 percent level. It is curious that $S0_2$ would be significant for chest pain but not for severe chest pain. However, the severe chest pain variable is described as one that lasts one half hour or longer which may not adequately reflect the potential chronic effects of either $S0_2$ or TSp.

TABLE 6.5 ALTERNATIVE ORDINARY LEAST SQUARES RIXRESS1ONS WITH CNEST PAIN AS THE DEPENDENT VARIABLE. t Statistics ARE IN PARENTHESES

Dependent Variable and Regression #	Independent Variables DIET	SMKN	DRNK	LIQR	BRTH	EARN	TSPM	SO_2M	STFT	SUGR	FIBR	USFT
CHPN I	.0512 (4.285)	.0273 (2.787)	.0080 (.589)	.00003 (3.344)	-.00002 (-.010)	-.0013 (-2.947)	-.00002 (-.619)	.0001 (1.772)	-			
CHPN II	.0535 (4.452)	.0269 (2.743)	.0091 (.669)	.00003 (3.568)	-.0001 (-.036)	-.0012 (2.907)	-.00002 (-.580)	.0001 (1.756)	-	.0007 (1.948)	-	
CHPN III	.0540 (4.494)	.0277 (2.712)	.0088 (0.644)	.00003 (3.549)	.00001 (.004)	-.0012 (-2.775)	-.00002 (-.504)	-.0001 (1.719)	-.0071 (-. 768)	-.0018 (2.135)	-.0503 (-2.172)	.0059 (.644)
CHPN IV	.0522 (4.285)	.0280 (2.845)	.0070 (.516)	.00003 (3.576)	.0001 (.030)	-.0012 (-2.885)	-.00002 (-.539)	.0001 (1.699)	-			

	NTRS	IRON	CALC	THMN	NIAC	VITA	FATS	PROT	RIBF	CONSTANT	R^2	SSR DF
CHPN I										.2085 (5.108)	.0068	1515 7899
CHPN II										.169 (3.692)	.0073	1514 7898
CHPN III							-.0052 -1.25	.0043 (1.052)	-	.1659 (3. 000)	.0085	1512 7893
CHPN IV	-.1215 (-. 762)	.0188 (.209)	2.0581 (2.354)	-.0057 (-.050)	-.0237 (-.484)	-.1865 (-1.068)	-	-	.1232 (.646)	-.275 (-1.558)	.0084	1512 7892

Chest pain and severe chest pain symptoms are uniformly higher in individuals who have reported the necessity of dieting. This finding is collaborated by extensive medical research on the effect of excess weight on the liklihood of heart attacks and other cardiovascular problems (10).

In Table 7 are recorded a sample of the regression results obtained for the occurence of coronary heart attacks. The variable reflecting the need to diet is again highly positively significant. Smoking is less significant but still positive. Consumption of alcohol has a marginally significant effect while excessive drinking seems to have a negatively significant effect. Family earnings has the anticipated negative effect on the occurence of coronary heart attack. Of the air pollutant variables, TSP has a positive and highly significant impact on coronary heart attack. Alternatively, SO_2 is negatively related to coronary heart attack but the coefficient is only marginally significant. The consumption of more starches, fats, and nitrosamines has an apparent positive effect on heart attacks and protein a negative effect. Conceptually, from these regressions one could compare the effects of consumption of certain foods with suspended particulate as to relative effects on the prevalence of coronary heart attacks. That will not be done here because of the experimental nature of these results and the need for additional replication before the results can be accepted.

In Tables 8 and 9 are a sample of regression results for two respiratory symptoms, the presence of cough and shortness of breath. In both cases, TSP had a significant impact on their occurence, while SO_2 had a negative impact. For the presence of cough, smoking, liquor consumption, sugar intake, and TSP had highly significant positive effects. The need for dieting, family earnings, and fiber consumption had a negative impact. For shortness of breath, the need to diet, smoking, liquor consumption, and TSP had positive and significant effects on its incidence.

In Table 10, the "t" statistics are contrasted for the various symptoms and air pollutant variables. As was noted before, these do not vary greatly when dietary variables are included. Maximum average concentrations of TSP have a strong connection to the presence of three symptoms, coronary heart attack, cough, and with less significance, shortness of breath. Maximum average 24 hour concentration of SO_2 has a positive connection with the occurence of chest pain but a significant negative connection with coronary heart attack and shortness of breath. This anomolous result cannot be readily explained. However, SO_2 concentrations are higher in heavy manufacturing-industrial areas where workers doing physical labor may be in relatively better physical condition due to exercise. In consequence, the answer to the shortness of breath question might be biased since it references walking on level ground with other people. Healthier individuals resulting from physical exercise at work may not respond to the shortness of breath question even though there may be some respiratory impairment.

Table 11 presents elasticities of the incidence rate of a symptom with respect to air pollution. These elasticities represent point estimates of elasticity about the mean. They derived via the following formula

TABLE 6.6 ALTERNATIVE ORDINARY LEAST SQUARES REGRESSIONS WITH SEVERE CHEST PAIN AS THE DEPENDENT VARIABLE. t STATISTICS ARE IN PARENTHESES

Dependent Variable and Regression #	Independent Variables DIET	SMKN	DRNK	LIQR	BRTH	EARN	TSPM	SO_2M	STFT	SUGR	FIBR	USFT
SVCP I	.0290 (4.928)	.0120 (2.496)	-.0059 (-.878)	.000001 (.195)	-.0008 (-1.016)	-.0003 (-1.652)	.00001 (.360)	-.000005 (-.168)	-	-	-	
SVCP 11	.0290 (4.895)	.0121 (2.498)	-.0059 (-.882)	.000001 (.180)	-.0008 (-1.015)	-.0003 (-1.654)	.00001 (.358)	-.000005 (-.167)	-	-.00002 (-.1048)	-	
SVCP III	.0289 (4.865)	.0119 (2.457)	-.0059 (-.876)	.000001 (.167)	-.0008 (-.920)	-.0003 (-1.621)	.00001 (.384)	-.00001 (-.178)	-.0017 (-.366)	-.0002 (-.362)	-.0063 (-.555)	.0054 (1.210)
SVCP IV	.0296 (4-931)	.0120 (2.481)	-.0067 (-.995)	.000001 (.151)	-.0008 (-.961)	-.0003 (-1.629)	.00001 (.381)	-.00006 (-.200)	-			

	NTRS	IRON	CALC	THMN	N IAC	VITA	FATS	PROT	RIBF	CONSTANT	R^2	SSR DF
SVCP I										.0610 (3.035)	.0041	367 7899
SVCP II										.0621 (2.760)	.0041	367 7898
SVCP III							-.0034 (-1.680)	.0017 (.825)	-	.0662 (2.431)	.0046	367 7893
SVCP IV	-.0528 (-.673)	-.0104 (-.234)	.1766 (.410)	-.0315 (-.560)	.0107 (.445)	.0071 (.082)	-		-.0046 (-.049)	.0213 (.245)	.0048	368 7892

TABLE 6.7 ALTERNATIVE ORDINARY LEAST SQUARES WITH THE INCIDENCE OF CORONARY HEART ATTACK AS THE DEPENDENT VARIABLE. c STATISTICS ARE IN PARENTHESES

Dependent Variable and Regression #	Independent Variables											
	DIET	SMRN	DRNK	LIQR	BRTH	EARN	TSPM	SO_2M	STFT	SUGR	FIBR	USFT
CORN I	.1686 (7.248)	.0183 (.962)	-.0.427 (-1.617)	.00002 (1.287)	-.0156 (-4.758)	-.2740 (-3.316)	.0002 (2.763)	-.0001 (-1.267)	-			
CORN 11	. 164(I (7.017)	.0192 (1.009)	-.0450 (-1.701)	-00002 (1.006)	-.0155 (-4.722)	-.0028 (-3.3S9)	.0002 (2.721)	-.0001 (-1.250)	-	-.0016 (-2.095>	-	
CORN III	.1607 (6.847)	.0231 (1.211)	-.0513 (-1.939)	.00002 (1.066)	-.0158 (-4.802)	-.0028 (-3.354)	.0002 (2.688)	-.0001 (-1.297)	.0270 (1.500)	-.0019 (-1.165)	.0469 (1.042)	-.0433 (-.807)
CORN IV	.1551 (6. 55o)	.0250 (1.309)	-.0555 (-2.091)	.00002 (1.031)	-.0158 (-4.826)	-.0028 (-3.423)	.0002 (2.729)	-.0001 (1,335)	-			

Dependent Variable and Regression #	NTRS	IRON	CALC	THMN	NIAC	VITA	FATS	PROT	RIBF	CONSTANT	R^2	SSR DF
CORN I										.4529 (5.701)	.0120	5737 7899
COKN II										.5364 (6.036)	.0126	5734 7898
CORN 111			, -				.0103 (1.281)	-.0227 (-2.845)	-	.6911 (6.425)	.0143	5724 7893
CORN Iv	.5283 (1.70s)	.2718 (1.548)	1.2481 (.734)	-.4602 (-2.073)	-.1206 (-1.268)	-.4820 (-1.420)	-		-.5041 (-1.360)	.5089 (1.483)	.0149	5120

TABLE 6.8 ALTERNATIVE ORDINARY LEAST SQUARES REGRESSIONS WITH COUGH AS THE DEPENDENT VARIABLE. t STATISTICS ARE IN PARENTHESES

Dependent Variable and Regression #	Independent Variables											
	DIET	SMKN	DRNK	LIQR	BRTH	EARN	TSPM	SO*M	STFT	SUGR	FIBR	USFT
coat I	-.0166 (-1.873)	.0952 (13.873)	.0097 (.966)	.99996 (10.366)	-.0019 (-1.537)	-.0010 (-3.306)	.00006 (2.229)	-.00004 (-.896)	-	-	-	-
COGH 11	-.0163 (-1.832)	.0951 (13.106)	.0098 (.979)	.00006 (10.322)	-.0019 (-1.541)	-.0010 (-3.298)	.00006 (2.235)	-.00004 (-.8991)	-	.00009 (.340)	-	-
COGH 111	-.0158 (-1.771)	.0950 (13.021)	.0121 (1.201)	.00006 (10.348)	-.0018 (-1.438)	-.0010 (-3.096)	.00006 (2.361)	-.00004 (-.865)	-.0119 (-1.739)	.0013 (2.108)	-.0944 (-5.513)	.0051 (.758)
COGH Iv	-.0102 (-1.132)	.0931 (12.792)	.0103 (1.024)	.(30006 (10.308)	-.0018 (-1.474)	-.0009 (-3.012)	.00006 (2.350)	-.00004 (-.963)	-	-		

	NTRS	I RON	CALC	THMN	NIAC	VITA	FArs	PROT	RIBF	CONSTANT	R^2	SSR DF
COGH I										.0697 (2.305)	.0469	831 7899
COGH II										.0646 (1.908)	.0469	831 7898
COGH III							-.0024 (-.776)	.0053 (1.822)	-	.0571 (1.396)	.0512	828 7893
COGH Iv	-.1781 (-1.510)	-.2631 (-3.935)	-.3818 (-.590)	.0989 (1.17)	.1299 (3.585)	.4264 (3.298)	-		.2604 (1.844)	.0218 (.167)	.0490	829 7892

TABLE 6.9 ALTERNATIVE ORDINARY LEAST SQUARES REGRESSIONS WITH SHORTNESS OF BREATH AS THE DEPENDENT VARIABLE. t STATISTICS ARE IN PARENTHESES

Dependent Variable and Regression #	Independent Variables DIET	SMKN	DRNK	LIQR	BRTH	EARN	TSPM	SO_2M	ST FT	SUCR	FIBR	USFT
SHBR I	.0254 (2.240)	.0366 (3.937)	-.0421 (-3.276)	.00004 (4.8012)	-.0010 (-.613)	-.0013 (-3.168)	.00004 (1.162)	-.0001 (-2.195	-			
SHBR II	.0255 (2.234)	.0366 (3.935)	-.0421 (-3.270)	.00004 (4.768)	-.0010 (-.614)	-.0013 (-3.166)	.00004 (1.163)	-.0001 (-2.195)	-	.00002 (.054)	, -	
SHBR 111	.0256 (2.242)	.0350 (3.761)	-.0400 (-3.099)	.(30004 (4.689)	-.0007 (-.410)	-.0013 (-3.129)	.00004 (1.208)	-.0001 (-2.169)	-.0088 (-.998)	-.0003 (-.377)"	-.0411 (-1.873)	.1769 (2.044)
SHBR IV	.0295 (2.554)	.0343 (3.681)	-.039'1 (-3.038)	.00004 (4.741)	-.0007 (-.468)	-.0013 (-3.111)	.00004 (1.139)	-.0001 (-2.149	-			

Dependent Variable and Regression #	NTRS	IRON	CALC	THMN	NIAC	VITA	FATS	PROT	RIBF	CONSTANT	R^2	SSR DF
SHBR I										.0957 (2.470)	.0082	1363 7899
SHBR II										.09 (2.183)	.0082	1363 7898
SHBR 111							-.0123 (-3.140)	.0106 (2.714)	-	.0669 (1.276)	.0103	1360 7893
SHBR IV	-.4710 (-3.117)	-.0670 (-.782)	-.0283 (-1.31)1)	.1877 (1..734)	.0283 (.609)	.0931 (.562)	-		.2437 (1.348)	.2258 (1.350)	.0102	1361 7892

TABLE 6.10 "t" STATISTICS ON AIR POLLUTION COEFFICIENTS*, SELECTED REGRESSIONS, NAS TWINS DATA SET

Symptom	Maximum Average 24 hour Concentrations** SO_2	TSP
Cardiovascular System		
Chest Pain	1.77[a]	-0.62
Severe Chest Pain	-0.17	0.36
Coronary Heart Attack	-1.27[d]	2.76[b]
Respiratory System		
Cough	-0.90	2.23[c]
Shortness of Breath	-2.19[c]	1.16[d]

*With nearly 8,000 observations, the "t" distribution approaches the normal distribution

[a]Significant at the 96% confidence level

[b]Significant at the 99.6% confidence level

[c]Significant at the 98% confidence level

[d]Significant at the 87% confidence level

**The simple correlation coefficient between TSP and SO_2 is .22.

TABLE 6.11 ELASTICITIES OF THE INCIDENCE RATE OF A SYMPTOM WITH RESPECT TO AIR POLLUTION*

Dependent Variable	Independent Variables Maximum 24 hour average concentration	
	SO_2	TSP
Chest Pain	1.995	-1.04
Severe Chest Pain	-0.505	2.64
Coronary Heart Attack	-5.988	21.23
cough	-1.757	6.88
Shortness of Breath	-8.329	7.25

*Elasticities are derived from coefficients in Equation 1 for all dependent variables at the mean values of the dependent and independent variables. This number represents the percent change in the probability of occurance of the symptom depicted by the dependent variable as a result of a 1 percent change in the independent variable

$$\text{Elasticity} = \frac{\text{change in the dependent variable}}{\text{change in the independent variable}} \cdot \frac{\text{mean of the independent variable}}{\text{mean of the dependent variable}}$$

Note however, that the first ratio on the right hand side of the above formula is simply the coefficient in the regression equation on the variable in question. This procedure allows the researcher to express results in percentage terms which are independent of the units used.

Care should be taken when interpreting the elasticities presented in Table 11. One should remember that the dependent variable is a probability. In this context, elasticities in the table represent the percentage change in the probability of the occurence of the event depicted by the dependent variable as a result of a one percent change in the independent variable. For example, if the maximum 24 hour average concentration of total suspended particulate increases by one percent then there will be a corresponding 21.23 percent change in probability of experiencing a coronary heart attack. However, the initial probability of a coronary heart attack (incidence) was slightly less than 12 percent in the sample. These values range for $S0_2$ from a low of -8.33 to a high of 2.00. Corresponding values for TSP range from a low of -1.04 to a high of 21.23.

What can be tentatively concluded from these experimental results? First, there appears to be a statistically significant connection between ambient concentration of total suspended particulate and several disease symptoms associated with both the respiratory and cardiovascular systems. Of particular importance is a strong and apparently replicative relationship between the incidence of coronary heart attacks and TSP. The evidence on concentrations of SO_2 and symptoms is much less clear. $S0_2$ is positively related to the self-reported occurence of chest pain. However, from these statistical results, SO_2 is negatively related to severe chest pain, coronary heart attack, cough, and shortness of breath. These findings should raise questions as to the reliability of self-reported data and the appropriateness of the questions themselves across diverse socio-economic groups.

Finally, regression equations were run omitting in sequence the $S0_2$ variable or the TSP variable. The omission of one of the air pollution variables had no influence on the magnitude, sign, or statistical significance of the included air pollutant variable. This lead us to the conclusion that the estimates reported in Tables 5 through 9 are relatively robust with regard to magnitude and sign.

ECONOMIC COSTS FROM POLLUTION

Lave and Seskin's (11) famous study, published in 1977, was one of the first to examine the statistical relationship between air pollution and health. They estimated the effects of air pollution, i.e., sulfur oxides and total suspended particulate, on the total mortality rate. Using the foregone earnings approach, they estimated benefits of pollution abatement via the reduction in the mortality rate. Lave and Seskin did not incorporate the relationship between air pollution and symptoms. Their approach focused on the direct relationship between air pollution and death.

Several other studies have been performed which relate air pollution and health. Most of these studies use mortality or morbidity rates as measures of health. For example, Crocker et.al. (12), 1979, use the mortality rate for pneumonia, influenza, emphysema, bronchitis and early infant disease as well as the total mortality rate for dependent variables. They used a variety of different air pollution measures as explanatory variables, concluding that only particulate and sulfur dioxide have statistically significant effects on health. Liu/Yu (13), 1979, utilized total mortality rates and the morbidity rate for bronchitis as health measures. They chose to use total suspended particulate and sulfur dioxides as pollution variables. Using both linear and non-linear models, they found that $S0_2$ and TSP have significant effects upon mortality and morbidity rates.

In contrast, this study focuses on the chain of events which link air pollution to the cost of increased symptoms due to air pollution. This methodology represents a substantial departure from that used in earlier studies.

Regression analyses, reported on earlier , were used to analyze the relationship between the occurence of a symptom and the factors affecting the symptom. Therefore, where Lave and Seskin use the mortality rate as the dependent variable, this report uses the occurence of a symptom such as cough, shortness of breath, etc. Coefficients on the independent variables give the change in the probability of a symptom given a unit change in a factor affecting the symptom.

Emphasis of this study is placed on the derivation of estimates of the reduction in costs of disease incurred when air pollution is reduced. The first step in this analysis is to depict the relationship between symptoms and disease. Consider:

$$P(D) = P(S_y) \cdot P(D/S_y) \qquad (19)$$

where

$P(D)$ = the probability of occurance of disease,

$P(S_y)$ = probability of the occurance of a disease symptom, and

$P(D/S_y)$ = probability of the occurance of a disease given the presence of a symptom.

This equation illustrates that the probability of a disease occuring is the probability of having a symptom related to that disease multiplied by the probability of having the disease given that symptom.

As is evident from the analysis presented in the previous section, one of the determinants of disease symptoms is air pollution. Therefore, the probability of incuring a disease symptom, and the resultant probability of incuring the disease, is conditional upon a given level of air pollution. In this context equation (19) becomes:

$$P(D/\tilde{P}_o) = P(S_y/\tilde{P}_o).P\{D/Sy), \tag{20}$$

where P_o is some given level of air pollution. Note that the probability of disease given a symptom is assumed independent of the pollution level.

For a change in the given level of air pollution, we observe:

$$P(D/\tilde{P}_1) - P(D/\tilde{P}_o) = [P(S_y/\tilde{P}_1) - P(Sy/Fe)] \cdot P(D/S_y); \tag{21}$$

where $\tilde{P}_1$ is a new level of pollution. This implies that:

$$\Delta P(D/\Delta\tilde{P}) = \Delta P(S_y/\Delta\tilde{P}).P(D/S_y). \tag{22}$$

Equation (22) illustrates that, as a result of a change in the level of air pollution, the change in the probability of incurring a disease is equal to the change in the probability of incuring a symptom multiplied by the associated probability of incuring a disease given the symptom.

From this analysis, the expected cost of disease can be defined as:

$$E(C_D/\tilde{P}_o) = P(D/\tilde{P}_o).C\,D\,, \tag{23}$$

where C_D is the cost of disease. A change in the expected cost given a change in the pollution level is given by:

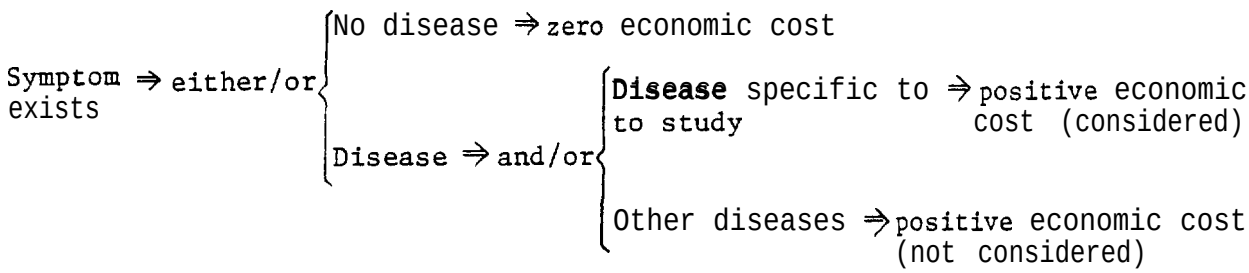
$$\Delta E(C_D/\Delta\tilde{P}) = \Delta P(D/\Delta\tilde{P}) \cdot C_D. \tag{24}$$

Substitution of equation (22) into equation (24) yields:

$$\Delta E(C_D/\Delta\tilde{P}) = CD \cdot \Delta P(S_y/\Delta\tilde{P}).P(D/S_y). \tag{25}$$

Equation (25) represents the change in the expected cost of disease given a unit change in the level of pollution for each symptom. The change in expected costs for each symptom can now be summed over the diseases to evaluate the total change in the expected cost of a symptom from a unit change in the pollution level.

Note that in the contex of the above analysis, an individual who has a disease symptom faces three possible states of the world. A symptom may exist and the individual has a disease or the symptom may exist without the corresponding presense of a disease. Further, since only certain rather specific diseases are considered in this analysis, it is possible that the individual who has a symptom does not have one of the diseases considered. The following diagram illustrates the possible situations:



Only the upper half of the bottom chain is considered in the definition of economic costs in this analysis. Therefore, this study only concentrates on the economic costs of a few diseases. Economic costs of other diseases are not considered.

One possible source of distortion in this analysis arises due to the fact that the economic costs incurred by a person who has several different diseases simultaneously is probably lower than the simple summation of economic costs from the individual diseases. In this aspect, medical costs are lower for an individual suffering from. several diseases than for several individuals suffering from one disease. This arises due to the fact that the same treatment procedures may apply to many diseases and that some costs, such as office calls, hospitalization, and loss of work time are relatively fixed once a disease is incurred. These costs tend to remain nearly the same whether one or several diseases are treated in the same individual.

Nine different diseases were used as representative of the circulatory and respiratory diseases which have these symptoms. Although there are many other diseases which are related to these symptoms, the inability to acquire data on alternate diseases prevented their use in this study.

The expected economic costs associated with these nine diseases were taken from alternative sources and adjusted to per case estimates (14)(15)(16). The total economic cost of a disease per case is the sum of the direct, indirect and expected mortality costs. Per case adjustments were made using morbidity and mortality rates. Table 12 presents the per case annual economic costs of each disease by type of expenditure. For example, the estimated expected average cost to an individual from having ichemic heart disease is $7,388.11 per year in 1981 dollars. Of this amount, $3,422 are direct expenditures which consist of hospital expenditures, nursing home fees and expenditures on physician services and prescriptions. Indirect costs, loss of work time due to illness is $3,720. The rest of the total expected cost is made up of the expected loss of earnings due to death. Expected lost earnings of the individual are discounted present values calculated with an 8 percent discount rate.

TABLE 6.12 ESTIMATED ANNUAL PER CASE EXPECTED COST OF DISEASES, BY TYPE OF DISEASE, IN 1969 DOLLARS

	Direct[a] cost	Indirect[b] cost	Expected[c] Mortality cost	Total[d] Expected cost	Number of Deaths/ Year	Prevalence/ Year (Thousands)
Respiratory Diseases[e]						
Chronic Bronchitis	$57 (154)	$30 (81)	$.90 (2.45)	$87.90 (237.45)	5,305	6,526
Bronchiectasis	198 (537)	60 (163)	.25 (.68)	258.25 (700.68)	1,476	116
Emphysema	130 (352)	344 (932)	3.82 (10.35)	477.82 (1294.35)	20,873	1,313
Chronic Interstitial Pneumonia	62 (168)	-	9.96 (26.98)	71.96 (194.98)	4,218	403
Heart Diseases[f]						
Ischemic Disease	1391 (3422)	1512 (3720)	100.05 (246.11)	2931.05 (7,388.11)	669,829	1,333
Rheumatic Fever and Rheumatic Heart Disease	291 (716)	407 (1001)	3.44 (8.47)	701.44 (1725.47)	15,432	327
Cardiomyopathy	15 (37)	96 (236)	3.66 (8.99)	114.66 (281.99)	17,753	1,560
Arrhythmias	325 (800)	139 (342)	1.49 (3.66)	465.49 (1145.66)	7,298	389
Cardiac Failure	2736 (6731)	418 (1028)	1.67 (4.12)	3155.67 (7763.12)	11,388	113

TABLE 12 (continued)

- = Insufficient data

[a]For heart disease direct costs = hospital expenditures + nursing home expenditures + expenditures on physician services. For respiratory disease direct costs = hospital expenditures + nursing home expenditures + expenditures on physician services + expenditures on prescriptions.

[b]Indirect cost = loss of earnings due to illness or disability.

'Expected mortality cost = expected loss of earnings due to death = (probability of death from disease) .(loss of earnings due to death). For respiratory disease a 6% discount rate is used, for heart disease an 8% discount rate is utilized.

[d]Expected total cost = direct + indirect cost + expected mortality cost.

'Heart disease data is in 1969 dollars and utilized 1969 and 1970 data. The figures in ( ) are adjusted to 1981 dollars.

[f]Respiratory data is in 1967 dollars and utilized 1967 and 1970 data. The figures in ( ) are adjusted for 1981 dollars.

References: 1. Acton, Jan Paul, "Measuring the Social Impact of Heart and Circulatory Disease Programs: Preliminary Framework and Estimates," Rand Corp. R-1697-NHLI, April 1975.

2. U.S. National Heart and Lung Institute, "Respiratory Diseases: Task Force Report on Problems, Research Approaches, Needs," DHEW Pub. No. (NIH) 76-432, pp. 205-243, October 1972.

3. Department of Health, Education and Welfare, National Center for Health Statistics, "Prevalence of Selected Chronic Respiratory Conditions," DHEW Pub. No. (HRA) 74-1511, Series 10, 84, 1970.

Expected values are a necessary component of the total cost of a disease since all individuals who have a disease do not necessarily die from the disease. This necessitates the use of an expected cost of mortality in the calculations. This number represents the loss of earnings due to death multiplied by the disease specific mortality rate. The mortality rate is the probability that an individual will die from the disease in question. Therefore, in this context the per case expected cost of disease becomes:

$$E(C_D) = d + i + E(m) \tag{26}$$

where

$E(C_D)$ = the expected cost of disease,

d = direct costs,

i = indirect costs, and

$E(m)$ = Probability of Death . Loss of Earnings due to Death = The Per Case Expected Cost of Death

Ideally, to depict the probability of death in this study, a mortality rate should be used which is conditional upon the presence of disease symptoms. However, since this information was unobtainable, per capita mortality rates derived for the society (of the U.S.) as a whole were used as a proxy. These rates are presented in the first column of Table 13.

Use of the societal mortality rate instead of a rate conditional on the existence of disease symptoms induced a downward bias to cost estimates. This is due to the fact that death rates due to disease are undoubtable higher in persons who already experience disease symptoms than in the society as a whole.

Note now that equation 25 must be modified to include the expected cost of disease. Equation 25 becomes:

$$\Delta E(C_D/\Delta\tilde{P}) = E(C_D).\Delta P(S_y/\tilde{P}).P(D/S_y) \tag{27}$$

Equation 27 forms the basis for derivation of cost savings due to reductions in the level of air pollution presented in this study. The first term on the left hand side, the per case expected cost of disease, is presented in Table 12. The second term, the change in the probability of incurring a disease symptom given a unit change in the level of air pollution, is simply the regression coefficient on air pollution variables which are presented in Section IV. The third and final term necessary to calculate the change in costs arising from a reduction in air pollution, the probability of disease given a symptom, is proxied in this analysis via the societal prevalence rate for the disease in question.

Again, as in the above discussion on mortality, use of the societal prevalence rate for a disease as a proxy for the incidence of that disease in individuals who already show evidence of symptoms will introduce a down-

TABLE 6.13 PER CAPITA PREVALENCE AND MORTALITY RATES OF SPECIFIC DISEASES IN THE UNITED STATES

	Mortality Rate	Prevalence Rate
Respiratory Diseases[a]		
Chronic Bronchitis	.00004	.03185
Bronchiectasis	.00001	.00057
Emphysema	.00018	.00641
Chronic Interstitial Pneumonia	.00004	.00197
Heart Diseases[b]		
Ischemic Disease	.00330	.00658
Rheumatic Fever and Rheumatic Heart Disease	.00007	.00161
Cardiomyopathy	.00009	.00769
Arrythmias	.00004	.00192
Cardiac Failure	.00006	.00056

[a]Based on number of deaths in 1967 and prevalence in 1970 from Table 12 and a U.S. population of 119,118,000 in 1967, U.S. Department of Commerce, Current Population Reports: Population Estimates and Projections, pg. 12, July 31, 1982, and a U.S. population 204,879,000 in 1970, Ibid., U.S. Department of Commerce, pg. 11, December 1972.

[b]Ibid., prevalence and deaths in 1969 from Table 12 and U.S. population of 202,677,000 in 1969, Ibid., U.S. Department of Commerce, pg. 11, December 1972.

ward bias to the results. This occurs due to the fact that, at the margin, the change in the probability of incurring a disease given a change in a symptom will be larger than the corresponding change in the incidence rate of that disease in the society as a whole. Societal prevalence rates for the nine diseases considered in this analysis are presented in the second column of Table 13.

Per capita estimates of the change in expected cost of disease given a unit change in the pollution level, derived via equation 27, are presented in the first column of Table 14. To derive these estimates, information from Table 12, Table 13 and the regression tables of Section IV are used. Note that these costs are presented by symptom and that they are adjusted to reflect 1981 dollars.

These results can be summed over diseases to yield per case estimates of the total cost of symptom given a unit change in air pollution. The last column of Table 14 presents these results. Note that not all symptoms apply to each disease and vice versa.

Table 15 presents estimates of cost of benefits in relation to unit changes in pollution levels. For extrapolative purposes, change in expected cost is assumed to be independent of the initial level of pollution. Intuitively, one would expect an increasing average relationship between the costs (benefits) incurred from a pollution increase (decrease) and the initial pollution level. This is illustrated garphically in Figure 5. If the initial level is P_0 and a change in the pollution level occurs bringing society to a level of P_1 the benefits received are B_1. Now if the initial level is $\tilde{P}_0$ and a reduction in pollution of the same amount as above occurs, $\Delta\tilde{P}$, the benefits received will be less than B_1 and are equal to $\tilde{B}_1$. However, it has been demonstrated that rather than increasing average benefits for increasing initial levels of pollution, there may be decreasing average benefits (17). Due to uncertainty surrounding the actual relationship, a linear relationship between pollution changes and economic costs is assumed to hold for purposes of extrapolating the results to larger pollution changes.

In order to derive estimates of total United States cost savings due to a reduction in air pollution, a 30 percent improvement in mean air quality is assumed. These results are presented in Table 16. Total cost savings are presented, by symptom, for males between the ages of 55 and 64 and for the total population in the United States. Male members of the U.S. population between 55 and 64 years of age most closely represent the twins sample as characterized by 1980 census data. A more proper characterization of the twins data set is to include all males 55 to 65 years of age in 1981. However, due to limitations in census data, this categorization is not possible. Approximately 10,178,000 males were in this age group in 1980. At that time, the total U.S. population was about 226,505,000.

A 30 percent reduction in average maximum 24 hour concentration of SO_2 and TSP implies that mean levels of SO_2 will be reduced by 14.88 $\mu g/m^3$ and TSP will be reduced by 38.86 $\mu g/m^3$. Therefore, total cost savings, per symptom, can be calculated via the following formula:

TABLE 6.14 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[a]

Symptom	Disease	$\Delta E(C_D/\Delta\tilde{P})$	Change in Total Cost of Symptom Given a Unit Change in the Pollution Level	
			TSP	SO_2
Cough	Chronic Bronchitis	.00045	.00391	
	Bronchiecstasis	.00002		
	Emphysema	.00050		
	Chronic Interstitial Pneumonia	.00001		
	Ischemic Heart Disease	.00292		
Shbr	Chronic Bronchitis	.00030	.00308[b]	
	Bronchiecstasis	.00002		
	Emphysema	.00033		
	Chronic Interstitial Pneumonia	.00002		
	Ischemic Heart Disease	.00195		
	Rheumatic Heart Disease	.00011		
	Cardiomyopathy	.00009		
	Arrhythmias	.00009		
	Coronary Heart Attack	.00017		
Chpn	Chronic Bronchitis	.00076		.00693
	Bronchiecstasis	.00004		
	Emphysema	.00083		
	Ischemic Heart Disease	.00486		
	Cardiomyopathy	.00022		
	Arrhythmias	.00022		

TABLE 14 (continued)

Svchpn	Ischemic Heart Disease	.00049	. 00053[b]
	Cardiac Failure	.00004	
Corn	Cardiac Failure	.00087	.00087

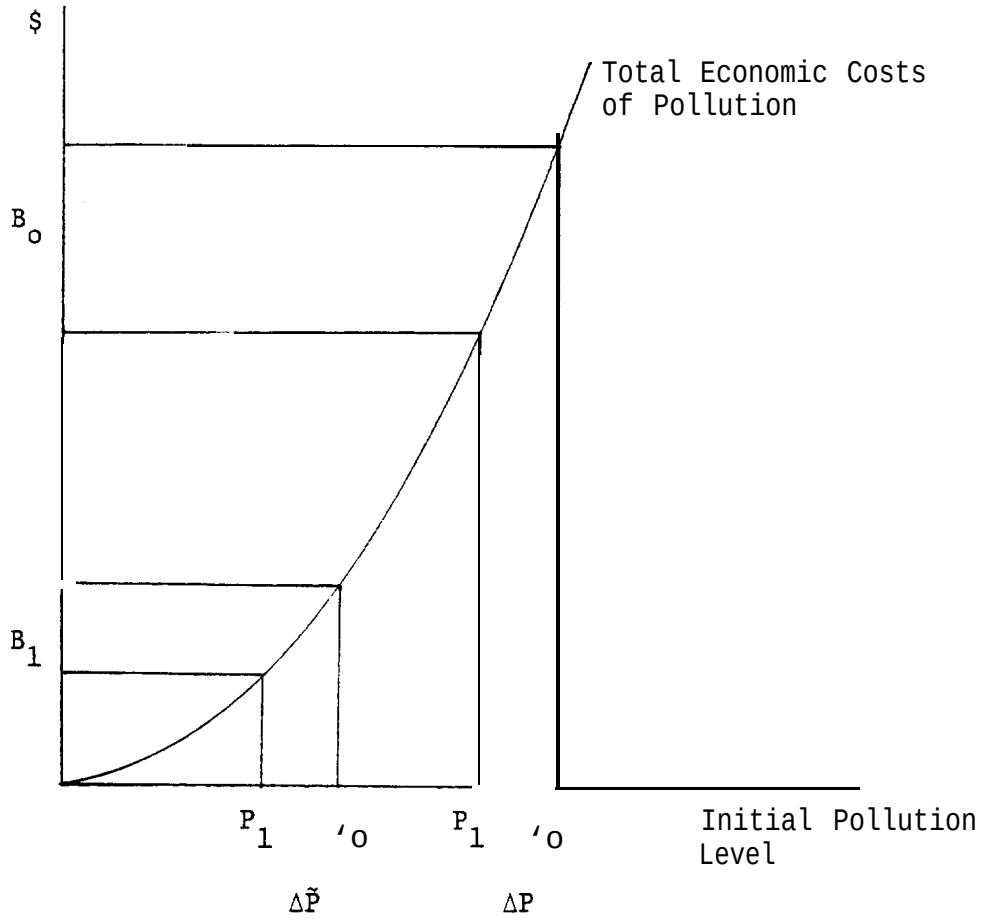
[a] values are reported only if the regression coefficient has a positive sign

[b] the coefficients used from the regression analysis to calculate these figures were not significant at the 90 percent level

TABLE 6.15 CHANGE IN PER CAPITA ANNUAL EXPECTED COST OF SYMPTOM GIVEN A CHANGE IN THE POLLUTION LEVEL

Symptom	Unit Change in the Pollution Level[a] $1\mu g/m^3$
Cough	.00391
Shortness of Breath	.00308
Chest Pain	.00693
Severe Chest Pain	.00053
Cardiac Failure	.00087

[a]TSP is used for all the symptoms except for chest pain where $S0_2$ is used.



Where $\Delta\tilde{P} = \Delta P$

Figure 6.5 Measuring benefits from pollution reduction assuming increasing costs of pollution

TABLE 6.16 TOTAL COST SAVINGS, BY SYMPTOM, FOR A 30 PERCENT IMPROVEMENT IN U.S. AIR QUALITY IN 1981 DOLLARS[a]

Symptom	Total for males between 55-64 years of age[b]	Total U.S. Population[b]
Cough	$1,546,000	$34,416,000
Shortness of Breath	1,218,000	27,110,000
Chest Pain[c]	1,050,000	23,357,000
Severe Chest Pain	210,000	4,665,000
Cardiac Failure	344,000	7,658,000
TOTAL	4,368,000	97,206,000

[a] mean values for SO_2 and TSP were used as initial values

[b] 1980 census of population data

[c] SO_2 is the air pollution variable used here and TSP is used for all other symptoms

$$\text{Total Cost Saving} = \text{Population} \cdot \text{Reduction in Air Pollution} \cdot \text{Per Case Cost of Symptom}$$

A 30 percent reduction in TSP is assumed for all symptoms except for chest pain where a 30 percent reduction in SO_2 is assumed.

Summation over the five symptoms yields an overall measure of the health benefits of air quality improvement. Note that for the age group nearest to the twins sample, total cost savings from disease is over $4 million. If these results are extrapolated to the entire U.S. population, a savings of nearly $100 million is incurred.

In order to compare this result to Ostro (1982) (18) and Crocker et.al. (1979) (19), it is necessary to exclude the cost savings arising from a reduction in SO_2 and only consider the costs savings arising from a reduction in total suspended particulate. Cost savings are reduced by $23,357,000 to $73,849,000 (in 1981 dollars) when only a 30 percent reduction in TSP is considered.

Ostro (1982) estimated that a 19 percent reduction in TSP will yield an urban benefit by reducing the number of work loss days by a range of 3 to 78 million. If a daily average wage of $46.00 is assumed for 1981, the range of damages in Ostro's analysis becomes $138 million to 3.588 billion.

Crocker et.al. (1979) analyzed the urban benefits of reduced mortality. Using the mean concentration of TSP in a sixty-city sample, they estimated the average reduction in risk of pneumonia mortality for a 60 percent reduction in particulate. Urban benefits of reduced mortality due to a 60 percent reduction in the level of total suspended particulate were estimated to be within a range of 5.4 to 16.7 billion dollars (adjusted to 1981 dollars).

In comparing the results presented in this paper to these other studies, one notes that the symptom sensitive analysis utilized here yields a lower bound. Only the lower end of Ostro's range is comparable with the results of this paper. Crocker et.al. estimates are much larger than the benefits estimated in either this study or Ostro's.

However, one can note that Ostro's results, which were calculated across all diseases, represent a marginal representation of work loss days. The indirect costs of disease presented in the d'Arge et.al. analysis were based on average work loss days due to a few specific diseases. In this aspect we would fully expect marginal work loss days to be larger than average work loss days because days lost increase as pollution increases.

Further, in considering the Crocker et.al. results, it must be realized that their results were based on the population as a whole while the d'Arge et.al. results were calibrated to a very specific sample of the population. At the time health statistics were collected for the twins data set, the group ranged in age from 41 to 51 years. In this context, the twins sample represented a fairly healthy segment of society. The Crocker et.al. sample included many older individuals whom we would expect would be more effected by

air pollution. Therefore, the Crocker et.al. result should exceed the d'Arge et.al. results in magnitude.

Finally, one should not forget the impact of the use of societal prevalence and death rates to proxy rates in individuals who exhibit disease symptoms in the d'Arge et.al. analysis. This phenomena will also result in the d'Arge et.al. results being lower bounds.

CONCLUSIONS

This analysis evaluated disease symptoms as related to smoking, consumption of alcohol, exposure to TSP and SO_2 outdoors, diet, age and earnings in 1973 as a proxy for socio-economic status. The study found that the only statistically significant relationship for air pollutants, which had the expected signs, were between TSP and cough and coronary heart attack and between SO_2 and chest pain. A slightly less significant relationship was found between TSP and shortness of breath.

The most significant "explanatory variables" for respiratory symptoms were dieting, smoking, alcohol consumption, socio-economic status, and air pollution. In this context, a positive relationship was found between shortness of breath and dieting, smoking, TSP concentrations, and one of the alcohol consumption variables. SO_2 and earnings were found to negatively effect shortness of breath. Dieting, age, earnings, and to a lesser extent SO_2 had negative effects on coughing while smoking, alcohol consumption and TSP had positive effects on the symptom.

The need to diet and smoking were consistently found to be positively correlated and economic status negatively correlated with cardiovascular system problems. Significant positive relationships between alcohol consumption and cardiovascular problems were found for chest pain and to a lesser extent coronary heart attack. Age was found to be negatively correlated with the occurance of all cardiovascular symptoms. However, a significant relationship between age and a symptom was only found for coronary heart attach. TSP was found to have a significantly positive effect on the incidence of coronary heart failure while SO_2 was found to positively effect chest pain. SO_2 was found to have a negatively significant effect on coronary heart attack. Finally, no air pollution variables were found to significantly influence severe chest pain. These findings suggest that the air pollution variables may be "masking" or replacing some other significant affects. Only similar analyses will perhaps lead to a net effect on ambient air quality on certain disease symptoms.

The list of symptoms were collected from the 1967-68 period while air pollution data were recorded for the year 1977, by zip code. Thus, only a weak inference can be made between air pollution common to times and symptoms. Because of time and manpower limitations, past air pollution data have not been included, inclusive of where the twin resided since 1945. Thus, unless the twin resided in the same place and there were no substantial changes in ambient air quality between the 1960's and late 1970's, the link between exposure and symptom can occur only be chance. Future research should center on more closely aligning symptoms with similar locations of exposure.

Evaluation of ordinary least squares and a more advanced technique of econometric analysis called "probit" yielded almost identical results except for a "scale" factor on the coefficients over at least fourty variants of

the preliminary model. This leads us to believe that OLS may be a reasonable technique to apply to more "robust" variables and theoretical systems.

Adequate variables measuring total inhalation of particulate, diet in terms of fat consumption, and "stress" variables have not been modelled. It is unlikely that current consumption of cigarettes, alcohol consumption as measured by a weighted sum of pure alcohol, or the need to diet, accurately reflect the impact on body processes. For example, a "heavy" smoker may have quit smoking in the early 1960's and yet retain some of the respiratory symptoms. Until these variables are adequately measured by complete exposure, it is unlikely that they will be useful for interpretation or prediction for policy purposes.

The effects of air pollution on health symptoms found in this study are roughly consistent with earlier work. However, with minor exception, all earlier sutdies focused on the effects of air pollution on mortality and morbidity. In four separate studies, Lave and Seskin (20)(21)(22)(23) McDonald/Schwing (24), Crocker (25), and Liu/Yu (26) all found partial linkages between air pollution and mortality and morbidity. Ostro (27) estimated the effects of total suspended particulate on work loss days. A comparison of the Ostro and Crocker et.al. results to the results presented in this study revealed that estimates presented in this study, as predicted, are of smaller magnitudes. Only Page (28) used a methodology remotely similar to the symptom-pollution relationships analyzed in this study. Page's measure of health effects was a self reported diary from 1,000 victims of respiratory illness as to whether they felt better, worse, or the same.

In order to derive total savings in health care costs, a 30 percent improvement in ambient air quality was assumed. The societal prevalence and death rates for nine diseases were used as proxys for the probability of incurring a disease or death given the presence of a symptom in the sample population. In this context, estimates of cost savings for a 30 percent reduction in maximum 24 hour ambient concentration of TSP and $S0_2$ was estimated to be over $4 million in males 55 to 65 years of age. Extrapolation of these savings to the total U.S. population yields an estimate of health cost savings of nearly $100 million.

APPENDIX 1

METHODOLOGY USED FOR FOOD CONVERSIONS

Table 17 presents the figures used to calculate the yearly consumption of different nutrients for the questionnaire respondents. In order to calculate Table 17, several assumptions were made on the serving sizes, given a questionnaire response. These assumptions, along with the figures in Table 19 were used to estimate Table 17. Figures in Table 19 were gathered from alternate sources (29)(30)(31)(32).

The following procedure was used to calculate nutrients ingested per year from consuming pasteries and candies:

(1) if more than one response was given the sample was deleted, and

(2) if only one response was given then the following was assumed:

Response	Assumption
0 never	0 serving/day
1 several times a day	3 servings/day
3 once a day	1 serving/day
5 less often	.5 serving/day

Nutrients in pork, frankfurters, beef, cereal, eggs, fish, vegetables and fruit were determined via the following procedure.

(1) if more than one response was given the sample was deleted, and

(2) if only one response was given then the following was assumed:

Response	Assumption
0 never	0 servings/day
1 daily	1 serving/day
3 once or twice/week	6 servings/month
5 once or twice/week	1.5 servings/month
7 less often	6 servings/year

For example, to determine the grams of protein consumed from eating a serving of frankfurters daily, multiply the 7 grams/day from Table 17 by 365 days in the year, i.e.,

$$7 \text{ gr/day} \cdot 365 \text{ days/year} = 2555 \text{ gr/year}$$

which gives the yearly consumption of protein from consuming frankfurters daily. If the respondent answered that he consumed frankfurters once or twice a month, it was assumed they consumed 1.5 servings per month. Therefore

the equation to calculate the grams of protein ingested in a year is

1.5 servings/month.7 gr/serving .12 months/year = 126 gr/year.

The yearly consumption of a nutrient for each respondent may be calculated by summing over the types of food for each nutrient. The yearly figures were used in the regression analysis to determine the importance of these nutrients to different symptoms reported.

TABLE 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[1]

Nutrient / Type of Food/var #	Protein (gm)	Fats (gm)	Fatty Acids uns (gm)	Fatty Acids sat (gm)	Carbohydrates sugar (gm)	Carbohydrates fiber (gm)	Vit. A (iu)	Ribo-flavin (mg)	Niacin (mg)	Thiamin (mg)	Calcium (mg)	Iron (mg)
Pasteries[2]												
(51) 1	5475	16425	10950	5475	32850	0	219000	109.5	547.5	54.75	36135	657
3	1825	5475	3650	1825	10950	0	73000	36.5	182.8	18.25	12045	219
5	912.5	2737.5	1825	912.5	5475	0	36503.65	18.25	91.25	9.13	6025.5	109.5
Candy[2]												
(52) 1	17520	19710	3285	5475	35040	0	175200	219	219	43.8	328.5	657
3	1460	6570	1095	1825	11680	0	58400	73	73	14.6	109.5	219
5	730	3285	547.5	912.5	5840	0	29200	36.5	36.5	7.3	54.75	109.5
Bread White (53)	$X_3 \cdot 730$	$X_3 \cdot 365$	na	na	$X_3 \cdot 5091.75$	$X_3 \cdot 18.25$	-	$X_3 \cdot 21.9$	$X_3 \cdot 255.5$	$X_3 \cdot 25.55$	$X_3 \cdot 8760$	$X_3 \cdot 255.5$
Whole Milk (54)	$X_4 \cdot 3285$	$X_4 \cdot 3285$	$X_4 \cdot 1095$	$X_4 \cdot 1825$	$X_4 \cdot 4380$	0	$X_4 \cdot 127750$	$X_4 \cdot 149.65$	$X_4 \cdot 73$	$X_4 \cdot 25.55$	$X_4 \cdot 105120$	$X_4 \cdot 36.5$
Skim Milk (55)	$X_5 \cdot 3285$	-	-	-	$X_5 \cdot 4380$	0	$X_5 \cdot 3650$	$X_5 \cdot .44$	$X_5 \cdot .73$	$X_5 \cdot 32.85$	$X_5 \cdot 108040$	$X_5 \cdot 36.5$
Coffee (56)	$X_6 \cdot 109.5$	$X_6 \cdot 36.5$	na	na	$X_6 \cdot 292$	0	0	$X_6 \cdot 36.5$	$X_6 \cdot 328.5$	$X_6 \cdot 3.65$	$X_6 \cdot 1679$	$X_6 \cdot 83.95$
Coffee w/ tsp. sugar (57)	$X_7 \cdot 109.5$	$X_7 \cdot 36.5$	na	na	$X_7 \cdot 4307$	0	0	$X_7 \cdot 36.5$	$X_7 \cdot 328.5$	$X_7 \cdot 3.65$	$X_7 \cdot 1679$	$X_7 \cdot 83.95$
Tea (58)	$X_8 \cdot 36.5$	0	na	na	$X_8 \cdot 328.5$	0	0	$X_8 \cdot 14.6$	$X_8 \cdot 328.5$	0	$X_8 \cdot 1825$	$X_8 \cdot 73$
Tea w/ tsp. sugar (59)	$X_9 \cdot 36.5$	0	na	na	$X_9 \cdot 4343.5$	0	0	$X_9 \cdot 14.6$	$X_9 \cdot 328.5$	0	$X_9 \cdot 1825$	$X_9 \cdot 73$

TABLE 15 (continued)

Pork[3]												
(60) 1	7300	8760	4380	3825	0	0	o	80.30	1715.50	284.70	3825	985.5
3	1440	864	432	324	0	0	0	7.92	169.20	28.80	324	97.2
5	360	432	216	162	0	0	0	3.96	84.60	14.04	162	48.6
7	120	144	72	54	0	0	0	1.32	28.20	4.68	54	16.2
Frankfurters[3]												
(61) 1	2555	5475	na	na	365	0	na	40.15	511. C	292	1095	292
3	252	540	na	na	36	0	na	3.96	50.4	28.8	108	28.8
5	126	270	na	na	18	0	na	1.98	25.2	14.4	54	14.4
7	42	90	na	na	6	0	na	.66	8.4	4.8	18	4.8
Beef[3]												
(62) 1	7300	9855	4745	4745	0	0	18250	58.40	1460	18.25	3825	912.5
3	1440	972	468	468	0	0	1800	5.76	144	1.8	324	90
5	360	486	234	234	0	0	900	2.88	72	.9	162	45
7	120	162	78	78	0	0	300	.96	24	.3	54	15
Cereal[3]												
(63) 1	730	-	na	na	7665	0	0	7.38	182.5	40.15	1460	146
3	72	-	na	na	756	0	0	.72	18.0	3.96	144	144
5	36	-	na	na	378	0	0	.36	9.0	1.98	72	7.2
7	12		na	na	1.26	0	0	.12	3.0	.66	24	2.4
Eggs[3]												
(64) 1	4380	4380	2190	14 60			430700	109.5	-	36.50	19710	803
3	432	432	216	144		-	42480	10.8	-	3.6	1944	79.2
5	216	216	108	72			21240	5.4	-	1.8	972	39.6
7	72	72	36	24			7080	1.8	-	.6	324	13.2
Fish[3]												
(65) 1	6205	1825	365	365	1825	o	na	21.9	985.5	10.95	12410	365
3	612	180	36	36	180	0	na	2.16	97.2	1.08	1224	36
5	306	90	18	18	90	0	na	1.08	48.6	.54	612	18
7	102	30	6	6	30	0	na	.36	16.2	.18	204	6
Vegetables[3]												
(66) 1	1095		na	na	8103	292		18.25	730	47.45	3650	292
3	108		na	na	799.2	28.8		1.80	72	4.68	360	28.8
5	54		na	na	399.6	14.4		.90	36	2.34	180	14.4
7	18	-	na	na	13).2	4.8		.30	12	.78	60	4.8

TABLE 6.17 (continued)

Fruit[3]												
((,7) 1	-	-	na	na	5840	730	18250	7.30	36.5	14.60	2920	146
3	-	-	na	na	576	72	1800	.72	3.6	1.44	288	14.4
5	-	-	na	na	288	36	900	.36	1.8	.72	144	7.2
7	-	-	na	na	96	12	300	.12	.6	.24	48	2.4

Footnotes: (1) There are two types of figures here, Var. 51, 52 and 60-67 already have the questionnaire response included within the calculation and only need to he identified by response. Var. 53-59 do not have response included in the calculation and therefore Lhe coefficient must be multiplied by Lhe response.

(2) If more than two responses where given on Lhe questionnaire then these samples were deleted. If this is not the case, the following was assumed.

Response	Assumption
0 never	1) servings/day
1 several times a day	3 servings/day
3 once a day	1 servings/day
5 less often	.5 servings/day

(3) Again if more than one response was given the sample was dropped and the following assumptions were made for the samples used.

Response	Assumption
0 never	0 servings/day
1 daily	1 servings/day
3 once or twice a week	6 servings /month
5 once or twice a month	1.5 servings/month
7 less often	6 servings/year

Notes: na: suitable data was notavailable but I he nu trient is suspected LO be present
- : only a trace has been detected
0 : the nutrient is not present and is I1oL suspected to be so

References: 1. Hamilton, E.M. and E. Whitney, Nutrition: Concepts and Controversy
2. Nutrition Search Co., Nutrition Almanac, McGraw Hill Book CO., 1975
3. National Dairy Council, Guide LO Good Eating, 1980

TABLE 6.18 FIGURES USED TO CALCULATE YEARLY CONSUMPTION OF NITROSAMINES BY QUESTIONNAIRE RESPONDENTS BY TYPE OF FOOD CONSUMED AND QUESTIONNAIRE RESPONSE

Type of Food	(Var. #)	Response	Nitrosamines (µg)
Pork	60	1	31.03
		3	3.06
		5	1.53
		7	.51
Frankfurters	61	1	224.84
		3	22.18
		5	11.09
		7	3.70
Beef	62	1	na
		3	na
		5	na
		7	na
Fish	65	1	31.03
		3	3.06
		5	1.53
		7	.51

Note: Minimum values are used here

References: Unpublished manuscript by Ron Shank for EPA Nitrates report

TABLE 6.19 LEVELS OF NUTRIENTS AND NITROSAMINES PER SERVING BY TYPE OF FOOD

Nutrients / Type of Food	Serving	Protein (gm)	Fats (gm)	Fatty Acid uns (gm)	Fatty Acid sat gm)	Carbohydrates sugar (gm)	Carbohydrates fiber (gm)	Vit. A (iu)	Riboflavin (mg)	Niacin (mg)	Thiamin (gin)	Calcium (ma)	Iron (mg)	Nitrosamines (µg)
Pasteries (51)	1 avg+	5	15	10	5	30	0	200	.10	.5	.05	33	6	0
Candy Milk Choc (52)	2 oz+ bar	4	18	3	5	32	0	160	.2	.2	.04	30	.6	0
Bread White (53)	1 slice* 22 slice/ loaf	2	1	na	na	13.95	.05		.06	.7	.07	24	.7	0
Whole Milk (54)	1 glass*	9	9	3	5	12	0	350	.41	.2	.07	288	.1	0
Skim Milk (55)	1 glass*	9	-	na	na	12	0	10	.44	.2	.09	296	.1	0
Cof fee (56)	1 cup*	.3	.1	na	na	.8	0	0	.01	.9	.01	4.6	.23	0
Coffee WI tsp. sugar (57)	1 cup w/1 tsp. sugar	.3	.1	oa	na	11.8	0	0	.01	.9	.01	4.6	.23	0
Tea (58)	1 cup*	.1	-	na	na	.9	0	0	.04	.1	0	5.0	.20	0
Tea w/ tsp. sugar (58)	1 cup w/1 tsp. sugar	.1	-	na	na	11.9	0	0	.04	.1	0	5.0	.20	0
Pork (60)	3 oz≠	20	24	12	9	0	0	0	.22	4.7	.78	9	2.7	.085
Beef (62)	3 Oz≠	20	27	13	13	0	0	50	.16	4.0	.05	9	2.5	na

TABLE 17 (continued)

Frankfurters (61)	2 oz≠	7	15	na	na	1	0	na	.11	1.4	.8	3	.8	.616
Cereal Cornflakes (63)	1 Cnp no sugar	2		na	na	21	0	0	.02	.5	.11	4	.4	0
Eggs (64)	2	12	12	6	4		0	1180	.3		.10	54	!.2	0
Fish Haddock (65)	3 Oz≠	17	5	3	1	5	0	na	.06	2.7	.03	31	1	.085
Vegetables (66)	1 cup	3		na	na	22.2	.8		.05	2.0	.13	10	.8	0
Fruit-apple (67)	1 med			na	na	16	2	50	.02	.1	.04	8	.4	.0

Footnotes: (1) all figures came from reference (1) except for those which came from reference (2).

Notes: * - These foods are measured in same manner as in questionnaire
≠ - Daily recommended servings are not used here as both references 1 and 2 used 3 oz. as an average serving
+ - Given there are no dally recommended servings for these variables. We assumed the average serving of pastery as 1 and an average serving of candy as a candy bar

preferences:
1. Hamilton, E.M. and E. Whitney, Nutrition: Concepts and Controversy, West publishing Co., St. Paul, Minnesota, 1979.
2. National Dairy Council, Guide to Good Eating, 1980.
3. Nutrition Search Co., Nutrition Almanac, McGraw-Hill Book Co., 1975.
4. Shank, R., unpublished manuscript for EPA Nitrate's Report, ch. 8, 1977.

REFERENCES

1. Bartel, A., and P. Taubman, "Health and Labor Market Success: The Role of Various Disease s," Review of Economics and Statistics 59, February 1979, H. Tuft, "The Impact of Poor Health on Earnings," Review of Economics and Statistics 57, February 1975.

2. Crocker, T. et.al., Studies on the Economics of Epidemiology, U.S. Environmental Protection Agency, technical report, volume 1, 1979.

3. Miller, M., and L. Leaven, Anatomy and Physiology, 16th edition, Macmillan Publishing Co., Inc., 1972. Especially Chapter 19.

4. A complete discussion of the NAS-NRC Twin Registry can be found in Zdenek Hrubec and James V. Neel, "The National Academy of Sciences - National Research Council Twin Registry: Ten Years of Operation," in Twin Research: Biology and Epidemiology, New York: Alan R. Teis, 1978.

5. Since it can be expected that the "average" health status of those serving in the armed forces is higher than those serving and not serving in the same age group, the sample is likely to have a higher health status than the U.S. population.

6. Zygosity is classified as either monozygotic (MZ) for identical twins and dizygotic (DZ) for fraternal twins.

7. U.S. Environmental Protection Agency, SAROAD: Information, Research Triangle Park, North Carolina, February 1979.

8. See the various reports from the United States Surgen General's office on the effects of smoking and health.

9* See, for example, Gould, Lawrence, "Cardiac Effects of Alcohol," American Heart Journal, volume 74, January-March 1970.

10. Clayton, D.G., J.W. Marr, and J.N. Morris, "Diet and Heart: A postscript," British Medical Journal 6096, November 1977, pp. 1307-1314.

11. Lave, L.B., and Eugene Seskin, Air Pollution and Human Health, Baltimore: Resources for the Future, 1977.

12. Crocker, T.D., W. Schulze, S. Ben-David, and A.V. Kneese, Methods Development for Assessing Air Pollution Control Benefits, Volume 1: Experiments in the Economics of Air Pollution Epidemiology, EPA-600/5-79-001a, Environmental Protection Agency, Research Triangle Park, N.C., 1979.

13. Liu, B.C., and E.S.A. Yu, Air Pollution Damage Functions and Regional Damage Estimates, Technomic Publishing Co., 1979.

14. Acton, J.P., "Measuring the Social Impact of Heart and Circulatory Disease Programs: Preliminary Framework and Estimating," Rand Corp., R-1697-NHLI, April 1975.

15. U.S. National Heart and Lung Institute, "Respiratory Diseases: Task Force Report on Problems, Research Approaches, Needs," DHEW Pub. No. (NIH) 76-432, October 1972, pp. 205-243.

16. Department of Health, Education and Welfare, National Center for Health Statistics, "Prevlaence of Selected Chronic Respiratory Conditions," DHEW Pub. No. (HRA) 74-1511, Series 10, 84, 1970.

17. Church, A.M., R.G. Cummings, and A.F. Mehr, "Respiratory Health Effects from Air Pollution: An Overview," University of New Mexico, draft report, 1981.

18. Ostro, Bart D., "The Effects of Air Pollution on Work Loss and Morbidity," submitted to the Journal of Environmental Economics and Management, 1982.

19. Ibid, Crocker et.al.

20. Lave, L.B., and E.P. Seskin, "Air Pollution and Human Health," Science, volume 169, 1970, p. 723.

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

22. Lave, L.B., and E.P. Seskin, "Does Air Pollution Cause Mortality?," in Proceedings of the Fourth Symposium on Statistics and the Environment, Washington, American Statistical Association, 1977, p. 25.

23. Lave, L.B., and E.P. Seskin, Air Pollution and Human Health, Baltimore: Resources for the Future, 1977.

24. McDonald, G.C., and R.C. Schwing, "Instabilities of Regression Estimates Relating Air Pollution to Mortality," Technometrics, volume 15, 1973, p. 463.

25. Ibid, Crocker et.al.

26. Ibid, Liu/Yu.

27. Ostro, Bart D., "The Effects of Air Pollution on Work Loss and Morbidity," submitted to the Journal of Environmental Economics and Management, 1982.

28. Page, W.P., Economics of Involuntary Transfers, Springer Verlog, 1973.

29. Hamilton, E.M., and E. Whitney, Nutrition: Concepts and Controversy, West Publishing Co., St. Paul, Minnesota, 1979.

30. National Dairy Council, Guide to Good Eating, 1980.

31* Nutrition Search. Co., Nutrition Almanac, McGraw-Hill Book Co., 1975.

32. Shank, R., unpublished manuscript for U.S. Environmental Protection Agency Nitrates Report, chapter 8, 1977.

BIBLIOGRAPHY

Acton, J.P., "Measuring the Social Impact of Heart and Circulatory Disease Programs: Preliminary Framework and Estimating," Rand Corp, R-1697-NHLI, April 1975.

Bartel, A., and P. Taubman, "Health and Labor Market Success: The Role of Various Diseases," Review of Economics and Statistics 59, February 1979, H. Tuft, "The Impact of Poor Health on Earnings," Review of Economics and Statistics 57, February 1975.

Church, A.M., R.G. Cummings, and A.F. Mehr, "Respiratory Health Effects from Air Pollution: An Overview," University of New Mexico, draft report, 1981.

Clayton, D.G., J.W. Marr, and J.N. Morris, "Diet and Heart: A Postscript," British Medical Journal 6096, November 1977, pp. 1307-1314.

Crocker, T.D., W.D. Schulze, S. Ben-David, and A.V. Kneese, Methods Development for Assessing Air Pollution Control Benefits, Volume 1: Experiments in the Economics of Air Pollution Epidemiology, EPA-600/5-79-OOla, Environmental Protection Agency, Research Triangle Park, N.C., 1979.

Department of Health, Education and Welfare, National Center for Health Statistics, "Prevalence of Selected Chronic Respiratory Conditions," DHEW Pub. No. (HRA) 74-1511, Series 10, 84, 1970.

Department of Health, Education and Welfare, "Public Health Service, Smoking and Health: A Report of the Surgeon General, DHEW Pub No. (PHS) 79-50066, 1979.

Gould, L., "Cardiac Effects of Alcohol," American Heart Journal, volume 74, January-March 1970.

Hamilton, E.M., and E. Whitney, Nutrition: Concepts and Controversy, West Publishing Co., St. Paul, Minnesota, 1979.

Lave, L.B., and E.P. Seskin, "Air Pollution and Human Health," Science, volume 169, 1970, p. 723.

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

Lave, L.B., and E.P. Seskin, "Does Air Pollution Cause Mortality?," in Proceedings of the Fourth Symposium on Statistics and the Environment, Washington, American Statistical Association, 1977, p. 25.

Lave, L.B, and E.P. Seskin, Air Pollution and Human Health, Baltimore: Resources for the Future, 1977.

Liu, B.C., and E.S.A. Yu, Air Pollution Damage Functions and Regional Damage Estimates, Technomic Publishing Co., 1979.

McDonald, G.C., and R.C. Schwing, "Instabilities of Regression Estimates Relating Air Pollution to Mortality," Technometrics, volume 15, 1973, p. 463.

Miller, M., and L. Leaven, Anatomy and Physiology, 16th ediction, Macmillan Publishing Co., Inc., 1972. Especially chapter 19.

National Dairy Council, Guide to Good Eating, 1980.

Nutrition Search Co., Nutrition Almanac, McGraw-Hill Book Co., 1975.

Ostro, Bart D., "The Effects of Air Pollution on Work Loss and Morbidity," submitted to the Journal of Environmental Economics and Management, 1982.

Page, W.P., Economics of Involuntary Transfers, Springer Verlog, 1973.

Shank, R., unpublished manuscript for U.S. Environmental Protection Agency Nitrates Report, chapter 8, 1977.

Us. Environmental Protection Agency, SAROAD: Information, Research Triangle Park, North Carolina, February 1979.

Us. National Heart and Lung Institute, "Respiratory Diseases: Task Force on Problems, Research Approaches, Needs," DHEW Pub. No. (NIH) 76-432, October 1972, pp. 205-243.

Chapter VII

ANALYTICAL PRIORS AND THE SELECTION OF AN "IDEAL" AIR POLLUTION EPIDEMIOLOGY DATA SET

INTRODUCTION

Widespread concern with the health effects of economics benefits generated by air pollution control programs has provoked a number of statistical studies of the association between air pollution and health status. However, the appropriateness of methodology and accuracy of the results of these studies have been widely disputed. The purposes of this paper, therefore, are threefold. First, we examine the role of optimal decision rules in testing the validity of price information to produce "best" estimates of the human health losses attributable to air pollution and the economic valuation of these losses. Secondly, we examine the use of price-information decision rules in previous air pollution-human health studies. Finally, based on optimal decision rules, we summarize statistically accepted prior information about the elements of an "ideal" air pollution epidemiology data set.

Statistical estimation of the degradation of health due to air pollution and the economic valuation thereof requires the use of prior information decision rules in four principal areas: (1) model selection (e.g., simultaneous, recursive, errors in variables, or single equations); (2) choice of functional form and the dimension of the design matrix; (i.e., matrix of exogenous variables); (3) the choice of values assigned to each element of the design matrix, if under the control of the experimenter; and (4) choice of the density function of the dependent variable. Most statistical analysis involved regressing a dependent variable (usually mortality and morbidity rates on time-to-failure for a system) a set of covariates which have been postulated to explain the variation in the dependent variable. Imposing prior information through exact parametric restrictions (whether correct or not) reduces the variance of estimated parameters. However, if incorrect, the restrictions increase estimator bias. Thus , the use of prior information, which is always incorrect to same degree except by chance, necessarily involves a tradeoff between the bias and efficiency of estimated parameters.

We evaluate this tradeoff in terms of the risk, i.e., the expected loss associated with each estimated parameter, measuring loss as the squared error of each estimated parameter relative to its true value, risk equals the sum of estimated parameter variances and squared biases. Stated somewhat differently, the researcher must choose decision rules which maximize the net benefit from utilizing prior information, where the benefit of such action is the resulting variance reduction and the cost is the resulting increase in bias. He seeks a middle ground somewhere between the overly restrictive case (high bias, low variance) and the totally unrestrictive case (unbiased, high variance).

In seeking decision rules for imposing prior information which minimize risk, there are valuable guidelines for accepting or rejecting hypotheses of exact prior restrictions (the most common type) and inequality restrictions. Regardless of the correctness of equality restrictions the positive-part Stein-rule estimator introduced by Baranchik (1964) which possesses minimum risk compared to the unrestricted estimator or the pre-test estimator (based on the standard decision rule to accept or reject the null hypothesis at a pre-specified level of significance). In addition, if inequality restrictions are correct in sign, they always exhibit less risk than the unrestricted estimator [see Judge, et al., 1980].

Our general conclusion regarding previous analysis of the effects of air pollution on human health and the valuation of these impacts. is that the preponderence of attempts to impose prior information have failed to minimize risk. Weak priors have rarely been correctly (if at all) tested before being imposed, while other strong but untestable priors have been ignored. We also conclude that the ideal data set, based on optimal decision rules, is not comprised of an exhaustive set of explanatory variables, since this would lead to unacceptably large estimator variances. Conversely, the ideal data set does not consist of a design matrix which excludes potentially important explanatory variables previous to statistical testing. To the extent that magnitudes of explanatory variables are under the control of the experimenter, the values assigned to an ideal data set should minimize risk subject to a given experiment budget constraint. If variables are not under the experimenter's control, the composition of the design matrix should be determined by optimal statistical tests based on prior information. An ideal data set can only be defined in conjunction with such information.

The plan for the remainder of the paper is to examine optimal decision rules for the use of prior information in section II and, in light of this, provide a critical review of the epidemiological literature measuring the effects of air pollution on human mortality and morbidity in section 111. A similar review of the literature which attempts to value these adverse health

affects is presented in section IV. Based on statistically accepted priors, in section V we suggest superior data sets for potential analysis. Finally, conclusions about optimal use of prior information are drawn in section VI.

USE OF PRIOR INFORMATION

Statistical estimation of the effects of air pollution on human health is impossible without the use of some prior information. This may take the form of model selection, choice of function form and dimension of the design matrix, selection of the values of each element of the design matrix (for variables under control of the experimenter), and choice of the density function for the dependent variable. The imposition of prior restrictions in these areas leads to an increase in the efficiency of estimated parameters. However, if restrictions are incorrect, estimated parameters are biased [see Judge, et al., (1980, ch. 11)]. Thus, the inescapable act of imposing prior information requires that the econometric researcher walk a tightrope between efficiency, on the one hand, and bias, on the other.

We proceed, therefore, to seek information regarding the optimal use of prior information which minimizes risk. In the context of regression analysis, we first define loss as the cost incurred if our estimate of the true value of the parameter vector of β is $\hat{\beta}$. Adopting a squared error loss criterion, we may write loss as

$$L = (\underline{\beta}-\underline{\beta})'(\underline{\hat{\beta}}-\underline{\beta}), \quad (1)$$

involving the k-dimensional vectors β and f3. Risk is defined as the expected value of loss:

$$P = E\,[(\hat{6}-6)'(\hat{6}-6)], \quad (2)$$

which equals the sum of variances for each element of $\hat{\beta}$ plus the sum of squared biases for each element of $\hat{\beta}$. Our objective is to minimize the risk from imposing prior restrictions.

Choice of Functional Form and Dimension of the Desire Matrix

We first consider this objective for the choice of functional form and dimension of the design matrix within the context of the testing of nested hypotheses[1/] for a single equation regression model. Four types of prior information may be imposed: exact restrictions, stochastic restrictions, inequality restrictions, and prior density functions. We compare the risks of utilizing these types of prior information to that of the unrestricted estimator, the pre-test estimator, and the Stein-rule estimator. The pre-test

estimator is simply the standard nested hypotheses test procedure whereby the null hypothesis (generally $\beta = 0$) is accepted or respected based on some predetermined level of significance. One example of a pre-test estimator is accepting or rejecting nested models of the quadratic Box-Cox (1964) form based on pre-determined levels of the likelihood function. Restrictions on estimated parameters lead to the inverse semi-log, semi-log, translog, generalized linear, quadratic, generalized square root quadratic, and linear models. [See Berndt and Khaled (1979)]. Choice among these nested models is typically based on the likelihood ratio test statistic. Additional restrictions allow testing of hypotheses about consumer behavior (homotheticity, additivity, and symmetry) or cost, production and profit function (homotheticity, homogeneity).

Exact information is the most common type of prior restriction. If the exact prior information is correct, the restricted least squares estimates are "best" estimates (i.e., minimum variance, unbiased). Incorrect exact prior restrictions, however, lead to biased estimates, which have smaller variances than under the correct model. The risk for the restricted least-squares estimator increases monotonically and exceeds the constant risk of the unrestricted maximum likelihood estimator, (MLE) over a wide range of hypothesis error under the assumptions of the general linear model. Further, the pre-test estimator has greater risk than the MLE estimator over a wide range of hypothesis error and hence, is inadmissible under our risk function criterion.

Stein-rule estimators [see Judge, et al. (1980, pp. 432) and Judge and Bock (1978)] exhibit less risk over the entire parameter space than the unrestricted and restricted MLE estimators, and the pre-test estimator. The positive-part Stein-rule estimator involves testing the hypothesis that $\beta_o = 0$, where β_0 is a vector of K_2 parameters. If $u_{(k_2)}$, the value of the likelihood ratio statistic, is less than or equal to $C_{(k_2)}$, where

$$C_{o_{(k_2)}} < C^*_{(k_2)} < 2C_{o_{(k)_2}} \text{ and } C_{o_{(k_2)}} = F_{(k-2),(T-k)}/k(T-k+2),$$

where k is the total number of variables and T is the total number of observations, we exclude the k_2 variables from the model. Otherwise, we employ the Stein-rule to transform the unrestricted MLE estimates using $C_*(k_2)$ and $u_{(k_2)}$ [see Judge et al. (1980)]. A second type of prior

information involves the use of stochastic prior information. Restrictions are assumed to hold subject to a normally distributed random vector. The sampling results for this type of prior restriction are parallel to those for the equality restricted estimator [see Judge et al. (1980)]. Inequality constraints comprise a third type of restriction. The risk function for the inequality constraint (when the direction is correct) is less than or equal to that of the MLE over the whole range of the parameter space the risk of the inequality pretest estimator (again when the direction is correct) is less than that of the traditional pretest estimator over almost the entire parameter space [see Judge and Yancky (1978)]. This result, which is particularly powerful, has largely been ignored by applied econometricians. It implies that risk can be reduced, sometimes substantially, by imposing sign constraints on estimated coefficients, when these signs are prescribed by economic theory. Thus for example, estimated parameters in health effect-pollutant exposure studies should be constrained to be non-negative.

Finally, prior information may be imposed in regression analysis through Bayesian procedures [see Zellner (1971)] which require the selection of prior density functions. The Bayesian procedure, a systematic way of combining sample information with prior information expressed as a density function, minimizes average risk for correct prior densities. However, economists have made little use of this technique because of their general reluctance to specify and test prior densities. The use of priors in model selection is simply a generalization of the procedures of their use in determining functional form and dimension of the design matrix in a single equation context. The use of MLE estimators, pre-test estimators, and Stein-rule estimators to test the validity of restrictions on the parameters in a simultaneous system is totally analogous to their use in a single-equation model. Appropriate restrictions could yield a recursive systems, a system with unobservable variables (but identifiable equations), or a Zellner seemingly-unrelated equation system [see Zellner (1962)] as restrive forms of the general jointly dependent system. Full-in format estimates are consistent and asymptotically efficient. Although single-equation estimators of a simultaneous equation model are biased and inconsistent, they possess minimum variance. In small samples, their risk as measured by mean square error is generally much higher than that of the full- information methods, based on Monte Carlo experiments, even with extremes of multicollinearity, [see Atkinson (1978) and Johnston (1972)]. Thus, the modeller is well-advised to first estimate a simultaneous equation model, if justified by priors, and apply the positive-part Stein-rule estimator to test nested hypotheses on restricted coefficients. Even if incompletely specified, additional restrictions across equations on parameters and, possibly, disturbance covariances aid in identifying the response structure. In addition, when these same cross-equation restrictions are viewed as hypotheses, significance

tests may be used to assess the statistical validity of the model.

Unobserved variables are a special class of errors-in-measurement problems which include omitted explanatory variables, and simultaneous equation systems.

In the air pollution epidemiology literature, attempts to grapple with the measurement error issue have been few. Crocker-Schulze, et al. (1979) raise the simultaneity issue for both air pollution-induced mortality and morbidity. Page and Fellner (1978) employ factor and canonical correlation analysis to attack the unobserved variable problem with respect to air pollution-induced mortality. Otherwise, air pollution epidemiology research largely consists of a vast number or single-equation regressions. Let us briefly examine the relationship between simultaneous equations, unobserved variables, and errors-in-measurement and their impact on estimator risk with the following example. Following Weld and Jureen (1953), who argued that many simultaneous equation relationships involving jointly dependent variables are really recursive relationships, we trace the chain of events from pollutant exposure to behavior change in Figure 1. The outcome at each step in the sequence is conditioned by the outcome in the previous period. Thus, for example, pollution does not immediately affect self-reported disability but rather has a delayed effect via its impact upon metabolism and organ system functions. Consider the following expressions:

$$Y_1 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \varepsilon_1, \tag{3}$$

$$Y_2 = \beta_o + \beta_1 Y_1 + \beta_2 X_2 + \varepsilon_2, \tag{4}$$

where Y_1 and Y_2 are, respectively, organ system function and self-reported disability, X_1 is pollution, X_2 is a vector of the other predetermined variables, and the ε's are random disturbances. Given (3), estimating (4) is equivalent to estimating the reduced form equation,

$$Y_2 = \beta_o + \beta_1\alpha_0 + \beta_1\alpha_1 X_1 + \beta_1\alpha_2 X_2 + \beta_2 X_2 + \mu, \tag{5}$$

where $\mu = \varepsilon_2 + \beta_1\varepsilon_1$. If the contemporaneous disturbances in (3) and (4) are uncorrelated, single equation MLE of (3) and (4) are equivalent to full-information estimation of this system.

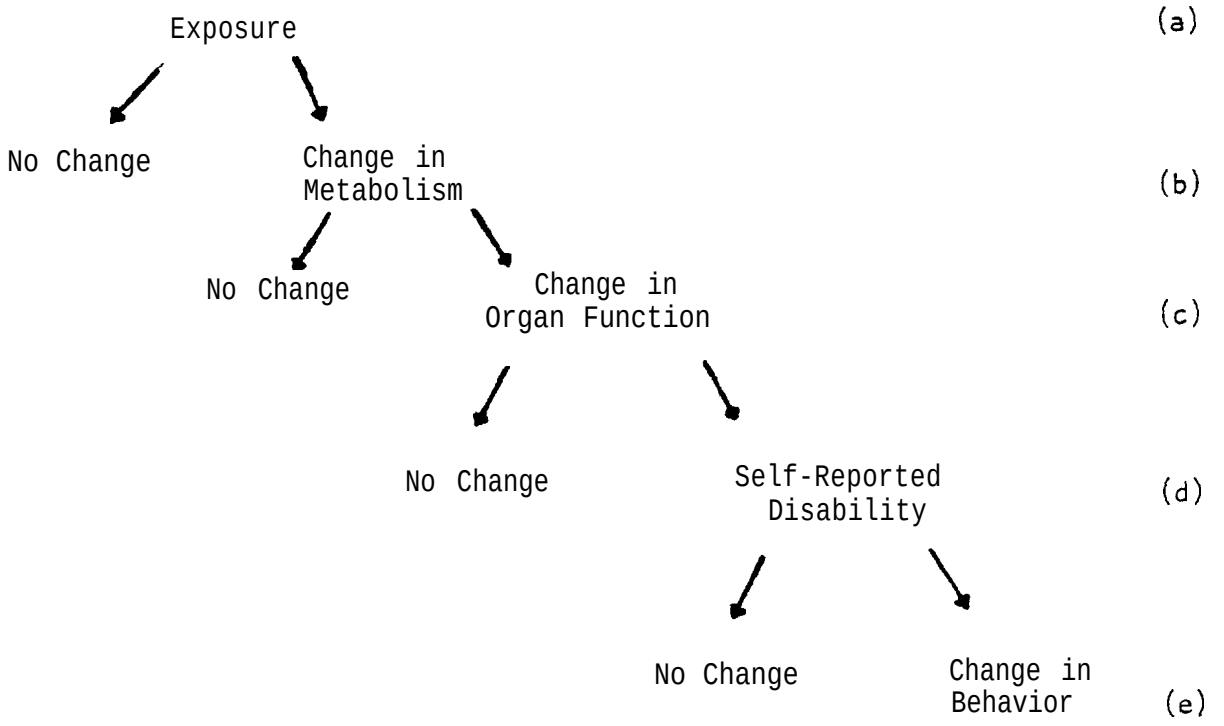
However, if Y_1 is unobservable, some investigators have simply estimated

$$'2 = 'o + \gamma_1 X_1 + \gamma_2 X_2 + \mu \tag{6}$$

Thus, if a MLE of (6) is to yield the same estimate of the impact of

FIGURE 7.1

A SCHEMATIC FOR AIR POLLUTION HEALTH EFFECTS



pollution, X_1, on self-reported disability, Y_2, as would a MLE of (4) given (3), γ_1 must equal $\beta_1 \alpha_1$. For this to occur, ξ_1 and ε_2 and the X's in (5) must be pair-wise uncorrelated [see Judge, et al. (1980, Chap. 13)]. Otherwise the estimate of y_1 will be biased and inefficient.

However, the random disturbances that influence organ system functions seem unlikely to be independent of factors affecting self-reported disabilities. For example, assuming that occupational exposures to toxics is not included among the explanatory variables of (3), and hence are part of the error, an exposure of this sort is likely to intensify the impact in (6) of any particular level of outdoor pollution upon self-reported disability. Instrumental variable methods, which involve the substitution into (4) of a proxy for Y_1 that is both highly correlated with it yet uncorrelated with ε_2, are available to overcome this problem. In the context of the structure represented by (3) and (4), it is not obvious what this proxy might be without additional prior information about (3). Further, use of a proxy in (6) would yield consistent but inefficient estimates of y_1. In short, whether an instrumental variable or a direct measure of X_1 is used, the power of the regression significance tests will most likely be reduced, requiring either a larger sample or more a priori information to maintain a given degree of test power.

Measures of the effective functioning of organ systems completely remove the necessity of wrestling with these particular estimation issues involving unobserved variables. This may be the reason that mortality rates and, more recently, time-to-system failure, have held great appeal as a measure of the health status of a population. Both the biomedical and the economic air pollution epidemiology literature would be considerably advanced through access to direct clinical measures of organ system functions or changes in metabolic processes.

Selection of Values of the Design Matrix

Having selected the appropriate model and the functional form and dimension of the design matrix, additional gains in efficiency can be achieved through the optimal choice of values of the design matrix. This includes both selection of the optimal values of the design variables under the control of the experimenter and the optimal number of observations of each selected value. Solution of this problem [see Figure 1, and Conlisk and Watts (1969)] involves minimizing an objective function, equal to a weighted function of the covariance matrix of the estimated parameters (where weights indicate the a priori importance attached to precise estimation of each variable) subject to a cost constraint on the experiment. The application of this technique to the creation of an epidemiological data base is straight forward. However,

again the estimator risk of this procedure depends on the risk associated with the exclusion of variables from the design matrix, the choice of functional form, and the choice of model to be estimated.

Choice of Density Function for Dependent Variable

The assumed density function of the dependent variable, and hence the error term, has been limited to the normal distribution for purposes of regression analysis throughout the economics literature. However, in many cases, the assumption of a normal density is unwarranted. When the dependent variable is a positive-valued variable representing either time-to-failure for a system or the mortality or morbidity rate for a specific population, previous empirical evidence yields strong priors which argue against the validity of a normal density. In fact, a substantial body of biomedical literature [see Kalbfleisch and Prentice (1980)] has made substantial use of non-normal models. The consequences of incorrectly assuming a normal density are estimator bias, since the parameters describing the likelihood function are incorrect, and possibly a loss in efficiency. Researchers in the biomedical area have adopted two principal models relying on non-normal density functions for the dependent variable in regression analysis. The first involves formulating a parametric regression model based on the generalized F distribution. Parametric restrictions on this distribution specialize it to the Weibull (which further specializes to the exponential), the generalized Gamma (which further specializes to the Gamma), the log logistic, and the log normal [see Kalbfleisch and Prentice (1980)]. Although hypothesis testing for nested densities has been carried out using MLE pre-test estimators, we recommend use of the positive-part Stein-rule estimator for the reasons discussed above. The second principal type of non-normal regression model is the partially parametric Cox (1972) proportional hazards (CPH) model or a non-proportional hazards generalization thereof. The CPH model is termed partially non-parametric because, with the introduction of appropriate parametric restrictions, it specializes to the Weibull and experimental regression models. In the case of a discrete dependent variable, the CPH model specializes to the logistic model. [See Kalbfleisch and Prentice, (1980, pp. 36-37)]. The CPH model has recently been applied to an increasingly wide number of regression problems attempting to explain system time-to-failure. The choice of a partially non-parametric model such as the CPH model in lieu of one of its nested counterparts (e.g., the Weibull or exponential regression models) is again based on minimum risk. Estimated parameters from the CPH model will have less bias that those from the nested models, but will be less efficient. However, Kalbfleisch and Prentice (1980) indicate that the CPH estimator possesses excellent relative asympotic efficiency as well as small sample efficiency compared to nested alterations. Thus, although the evidence regarding efficiency and risk is not

compete, the CPH model appears to afford a considerable increase in flexibility with little increase in risk. Additionally, it allows testing for and accepting its nested densities. The alternative of imposing one of the nested forms appears to offer little gain in efficiency at the risk of considerable increase in bias.

A CRITICAL REVIEW OF THE DOSE-RESPONSE LITERATURE

Over the past decade numerous studies of the economic value of the adverse health effects from air pollution have been carried out by economists and epidemiologists. The ultimate goal of these analyses has been the estimation of defensible functional relationships between dose and response, and then to estimate the resulting economic losses, so that marginal benefits of pollution reduction can be derived from them. The optimal level of pollution control can then be determined where the marginal benefit equals the marginal cost of additional pollution reduction. Recently, substantial controversy has developed over the adequacy and validity of certain methodological approaches and empirical results of studies quantifying dose-response relationships.

In general, there appears to be a minimal attempt in this literature to utilize prior information to formulate and test restrictions of the type previously discussed.

Although the health effects of air pollutants have long been studied in laboratories by toxicologists, there appears to be limited use of this information in non-laboratory studies by epidemiologists. Laboratory experiments on animals allow careful control of the level of individual pollutants, other covariates, and a detailed record of response. These studies, therefore, have been useful for identifying potential human health effects.

Laboratory experiments with human subjects avoid extrapolation from animal to man, but raise other concerns, such as ethical considerations and practical difficulties in studying long-term exposures. In addition, laboratory studies cannot duplicate the activity patterns and pollutant mixture experienced by free-living populations. Within these constraints, experiments involving human subjects can be conducted and used to establish levels at which adverse responses occur after short-term exposures.[27] Despite their limitations, much of what has been learned from laboratory studies could

be employed to provide structure for epidemiological studies. However, many epidemiological studies appear to ignore much of the toxicological literature by assuming linear dose-response functions, thereby failing to investigate possible synergistic effects among pollutants and other important personal factors[3] as well as more complex non-linear mathematical dose-response models based on non-normal distributions, which have been observed by toxicologists.[47]

Studies of occupational groups have been suggested as another source of information. Although such non-experimental studies may allow accurate estimates of exposure, the mix of pollutants and concentrations in workplaces is usually different than the mix in the general ambient air. Exposures are for only work hours rather than the entire day. Temperature and humidity conditions are also likely to differ in important ways from those experienced by the general population. The very young, elderly, and ill are not included. There is considerable selection by the employer and self-selection by the worker, so that those with current disease or those who are more sensitive or more susceptible are found among the employed less frequently than in the general population. Consequently, one cannot extrapolate from findings for occupational groups to the general population. On the other hand, if an association between an air pollutant and a health effect is found in an occupational setting, we would expect a greater association in the general population, if exposed to the same level of the particular pollutant.

In view of these limitations, most of the relevant information about the health effects of air pollutants at levels of exposure near present ambient conditions must come from observational studies of the general population. Here, too, there are limitations with respect to estimating exposure and measuring health effects. Uncontrolled variations in ambient pollution levels make it difficult to determine whether mean concentrations, peak concentration, the variance, or some other measure of air pollution concentration is the most important determinant of health. Additionally, pollution data are usually obtained from outdoor monitoring stations, but the actual exposure burden can vary greatly between individuals even living in the same neighborhood. Outdoor micrometeorology and indoor environment can significantly alter exposure [Benson, et al. , (1972)1 . This imprecision tends to bias estimated associations between air pollution and health effects toward zero. Moreover, health endpoints, including frequency of symptoms, lung function, hospital admissions, and cause of death also are measured with substantial variability. When an association between air pollution and health

is found, a high degree of collinearity between pollutants and the possibility of complex chemical interactions may make it very difficult to associate any health effect with a single pollutant.

Much of the recent work in air pollution epidemiology has focused upon estimation of a linear regression model based on the assumption of a normal error term, where a measure of the incidence of mortality or morbidity is regressed on air quality and other covariates. Many covariates are "personal" factors such as diet, smoking habits, exercise, medical care, age, sex, occupation, income, and genetic predisposition--while others are environmental factors-- such as quality of drinking water, toxic contamination, temperature, humidity, and exposure to allergins.

Many epidemiological studies originating in the biomedical disciplines and sanctified in existing Federal clean air legislation, assumes a positive level of air pollution or threshold below which no individual will suffer a decline in health status.[5/] However, this assumption is clearly a testable hypothesis. The first attempt to employ regression analysis to investigate the health effects of particulate and sulfate air pollution (i.e., principally stationary source pollution) at a national level without the presumption of a threshold was the pathbreaking effort of Lave and Seskin (1970). Using a cross-section of 114 U.S. metropolitan areas, they employed single equation, ordinary-least-squares methods to regress 1960 mortality rates upon ambient concentrations of sulfates and particulate, and other demographic and socio-economic variables. However, they maintained rather than tested the hypothesis that personal factors such as medical care, smoking, and ingestion of fat and alcohol were distributed independently of pollution levels. Thus, Lave-Seskin's analysis is immediately suspected of omitted variable bias, since there is substantial evidence that these factors synergistically interact with air pollution. They tentatively concluded that air pollution caused statistically significant health effects.

This original study has inspired a substantial number of similar studies, including the culminating effort of Lave and Seskin (1977). Included in this list are studies by Gregor (1977), Wyza (1978), Mendelssohn and Orcutt (1979), Seneca and Asch (1979), and Lipfert (1979) involving the mortality effects of sulfur oxides, sulfates, and particulate, and Schwing and McDonald (1976) involving the mortality effects of carbon monoxide, nitrogen dioxide, hydro-carbons, and photochemical oxidants. Studies of the morbidity effects of air pollutants include those by Jaksch (1973) and Seskin (1979). These mortality and morbidity, without exception, all have discerned a significant inverse association between mortality rates and one or more air pollutants, and in general these studies employ the model and functional form of Lave and Seskin. The results of these and more recent studies, which significantly question the

validity of the Lave-Seskin assumptions and results, are summarized in Table 1. V.K. Smith (1977), who used data for 50 U.S. metropolitan areas in 1968-1969, applied versions of the Ramsey (1969) tests for specification error in the general linear model to 36 different single equation specifications. These specifications were similar, and often identical, to those greeted with the most approval by Lave-Seskin, and others. None of the specifications could pass all of the Ramsey tests at the 10 percent level, although four passed all tests except that for non-normal errors which was rejected by all specifications. This result is particularly disturbing. Since Lave-Seskin estimated a linear single-equation model, the change or variable theorem indicates that the dependent variable, mortality rates, are also non-normally distributed. Thus, maximum likelihood techniques should have been employed to estimate a non-normal model, e.g., the Cox proportional hazard model or the Weibull or exponential regression models which are restricted cases thereof. This analysis could even be extended to include Bayesian prior distribution quality and other socio-economic and demographic variables.

Second, Thibodeau, et al. (1980) report on a limited reanalysis of the Lave and Seskin data. While they did notargue the existence of a health-pollution association, they questioned Lave and Seskin's methodology. In particular they found significant lack-of-fit and their reanalysis resulted in estimated effects which differed considerably from those reported by Lave and Seskin.

In a recent monograph, Crocker-Schulze, et al. (1979, pp. 24-71) analyzed 1970 mortality data from a cross-section of 60 cities while trying to correct for potential omitted independent variable and simultaneous equation misspecification. Adding measures of medical care, cigarette consumption, and diet to the single equation Lave-Seskin, specification, they found a nonstatistically significant effect of nitrogen dioxide, total suspended particulate, and sulfur dioxide upon the rate of total mortality,[67] in sharp contrast to the results of Lave and Seskin. Retaining the former variables and accounting for the plausible simultaneity between health status and medical care did nothing to improve the statistical significance of the three air pollution variables. On the presumption that these findings were sufficient to demonstrate the lack of robustness in the Lave-Seskin type results, the authors did not go on to account for the obvious simultaneity between median age (or incidence) and several other plausible sources of simultaneity.

The results of Crocker-Schulze et al. (1979), indicating that the Lave-Seskin type of analysis suffers from omitted variable bias, are given additional support by Graves, Krumm, and Violette (1979) who found significant synergisms between pollutant levels and personal factors in explaining mortality rates. Thus, Lave-Seskin should have tested rather than maintained the hypothesis

TABLE 7.1

A SUMMARY OF EPIDEMIOLOGICAL STUDIES OF AIR POLLUTION

The Effect of Air Pollution on Human Morbidity and Mortality

Mortality

Author	Model and Functional Form	Pollutants Used to Explain Level of Dependent Variable
Lave and Seskin (1970) (1977)	general linear model; linear regression	sulfur oxides and particulate
Crocker et al. (1970) model; linear regression of simultaneous equations	general linear sulfur dioxide and particulate[a]	nitrogen dioxide
Lipfert (1979a) model; linear regression	general linear particulate, and Sulfates[a]	sulfur dioxide
Gregor (1977) model; linear *regression*	general linear particulates[a]	sulfur dioxide
Seneca and Asch (1979) model; linear regression	general linear and sulfur dioxide	sum of particulate
Wyzga (1978) model; linear regression with lagged dependent variable	general linear	particulate

TABLE 7.1 (continued)

Mendelssohn and Orcutt (1979) regression	general linear model, linear	sulfates, carbon monoxide, and sulfur dioxide
Schwing and McDonald (1976)	general linear model; linear regression, ridge regression, and sign constrained least squares	hydrocarbons and nitrates[a]

Morbidity

Author	Model and Functional Form	Pollutants Used to Explain Level of Dependent Variable
Jaksch (1973)	general linear model; linear regression	particulates[a]
Crocker et al. (1979)	general linear model; linear regression and recursive linear regression	nitrogen dioxide sulfur dioxide, and particulate[a]
Graves and Krumm (1979)	general linear model; second order Taylor expansion	sulfur dioxide and particulate
Seskin (1979)	general linear model; linear regression	photochemical oxidant

a Indicates dependent variable explained by personal factors as well as air

that personal factors are independent of air pollution with the framework a simultaneous equation Box-Cox model.

The results obtained by V.K. Smith (1977), Thibodeau, et al. (1980), and Crocker-Schulze, et al. (1979) cast doubt upon the robustness of the Lave-Seskin, et al. estimates, in spite of the no-threshold perspective embodied in these estimates. These doubts are particularly bothersome when the results are extrapolated to project pollution regulation impacts. Nevertheless, before dismissing the hypothesis of an inverse relation between everyday air pollution levels and health states, it must be recognized that Lave-Seskin, et al., may have been asking more of their data than it was capable of giving.[7] Less than one in every 100 people dies in the U.S. each year. No biomedical authority asserts that air pollution is the dominant cause of the deaths that do occur. Many take the view that it is the direct cause of no more than a small fraction of these deaths, although they would agree that it may be quite important in intensifying predispositions toward mortality. However, the general properties of the underlying processes that encourage this predisposition are ill-understood. Thus, even with quite large samples, available estimation techniques and a priori knowledge may be inadequate for distinguishing the mortality effects of air pollution in a human population sample from a host of similar and plausible minor contributing factors.

The possible inadequacy of many available techniques for estimating the existence and/or magnitude of air pollutant-induced mortality applies with special force, given the data Lave-Seskin and their successors had to employ. Their work can be interpreted as an attempt at establishing the probability of a representative individual currently residing in a representative region dying in a given year from a geographically representative level of air pollution. Lave and Seskin justify their use of cross-section regional data on the grounds that these data reflect long-run adjustments by capturing response to pollution levels that have existed for long periods of time. Clearly, this assumption is questionable for many areas where pollution levels and populations at risk (due, e.g., to in and out migration) have changed over time. In addition, since they had no information about the distribution of covariates including air pollution across urban areas, the identifying variabilities of their samples were perhaps drastically reduced.[8] When this relatively low variability of the samples is coupled with what are probably substantial measurement errors in the air pollution variables, attempted corrections in model specification may serve only to misinform.

The preceding remarks lead us to three conclusions. First, given the biomedical and economic subtleties inherent in comprehending the etiologies of air pollution-induced mortality and morbidity, the estimates obtained from aggregated data used in the great bulk of extant studies are unlikely ever to

be sufficiently compelling to establish a consensus. Only when physiological models are coupled with observations on individuals can we expect compelling evidence. Second, statistical power should be substantially increased if research concentrates on morbidity rather than mortality. The frequency, and most likely the identifying variability, of morbidity data appears to be greater than that for mortality data by a factor of fifteen or twenty. Greater variability is also expected with more disaggregated data sets on mortality or mortality for the same reason. Finally, because one's health status is influenced by choices about lifestyles, environmental and occupational exposures to possible toxics, and other health-influencing factors, economics can provide a priori hypotheses and an analytical framework to lend additional structure to epidemiological investigations. The researcher can then further narrow the relationships with which observed real world outcomes can be compared. That is, the limited prior information from the existing epidemiological studies contribute something worthwhile to our goal of parsimonious data collection, but still confronts us with an enormously large parameter space, many elements of which could be insignificant for human health status. The more correct a priori information we can introduce to the problem, the greater the reduction in estimator risk. Given that health effect estimates are to be used for valuation assessments, efforts to reduce the severity of this tradeoff become particularly worthwhile.

A CRITICAL REVIEW OF THE VALUATION OF HEALTH EFFECTS LITERATURE

Economic Valuation of Mortality and Morbidity

Two principal methods of valuing mortality have been utilized in the empirical studies valuing human health. The first involves calculating the discounted present value of earnings lost due to mortality or morbidity [see Weisbrod (1971) and Cooper and Rice (1976)]. This is generally agreed to be an incorrect measure of the true value of mortality and morbidity, whose theoretically correct measure is either the willingness-to-pay to avoid mortality or the compensation required to voluntarily accept such adverse effects.[9/] At best, the discounted present value measure is a very limited estimate of the value of life (e.g., zero for the unemployed or retired) and does not allow for observed trade-offs in the job market between wages and risk of death or injury.

The second method of valuing mortality and morbidity involves estimating willingness-to-pay for risk reduction from: 1) surveys or questionnaires; 2) wage premiums for hazardous occupations; and 3) the cost and estimated effectiveness of safety devices. An individual's willingness-to-pay for a small reduction in the probability of death is generally extrapolated to

calculate the value of statistical life.

Two willingness-to-pay surveys have been conducted to estimate the value of life. Acton (1973) asked a sample of 37 Boston area residents to state their willingness to pay for emergency coronary care facilities which would reduce the probability of a fatal heart attack. From the responses, Acton estimated a value of life of less than $100,000 ($ 1978). Jones-Lee (1976) estimated a far higher value of life in excess of $6 million ($ 1978) for safer air travel, by asking travelers their willingness to pay higher fares to travel on airlines with lower probabilities of a fatal crash. However, difficulties in obtaining reliable estimates to theoretical questions arise because of incentives for strategic behavior, e.g., with public goods, and the limited ability of the individual to make an accurate determination of preferences in hypothetical situations. See Freeman (1979) for a discussion of attempts to overcome various types of strategic bias.

A more fruitful approach has been taken by a number of studies attempting to measure the value of life from data on wage differentials in hazardous occupations. Thaler and Rosen (1976) analyzed a sample of 900 individuals in 37 high-risk occupations taken from the records of the Survey of Economic Opportunity. They explained wage differentials among these occupations with: (1) the extent to which the risk of accidental death exceeded the expected average from statistical life tables; (2) regional and urban dummy variables; (3) demographic characteristics; and (4) job characteristic and occupational dummy variables. By extrapolating risk to zero, Thaler and Rosen calculated a value of life ranging from $273,000 to $508,000, with a best estimate of $391,000 ($ 1978). Using the same data on wages but different estimations of occupational risk, R.S. Smith (1976) obtained substantially higher estimates of the value of life, ranging from $2.2 million to $5.1 million ($ 1978). Finally, using a different data set, Viscusi (1976) obtained estimates ranging from $1.8 to $2.7 million ($ 1978) for blue-collar workers.

Three caveats must be applied to the use of these estimates. First, they represent the value of marginal changes in the probability of death extrapolated to a zero probability of death. If the marginal valuation of different probabilities varies significantly, this extrapolation may be highly biased. Secondly, the willingness to pay measured by these studies most likely is associated with accidental death and excludes the value of the disutility associated with the morbidity, pain, suffering with characterize fatal but chronic diseases such as cancer. Thus, these estimates may understate the willingness to pay by the general population. Finally, data on risk by occupation are not corrected for the fact that omitted personal characteristics are often associated with high risk jobs which account for non-job related deaths. Thus, a certain component of increased mortality

cannot be associated with a corresponding wage differential.

Studies estimating the willingness to pay by the general population for risk reduction as evidenced in consumer purchases of safety devices include those by Blomquist (1979) and Dardis (1980). Blomquist (1979) developed a simple life-cycle model of individual life-saving activity and estimates a value of life based on automobile seat belt use. Solution of his simple utility optimization model yields the first-order condition that the marginal value product of reduced mortality plus the marginal value product of reduced morbidity equals marginal cost. Blomquist then used probit analysis to explain the incidence of seat belt use with a set of demographic variables, length of work trip, speed limit, labor wealth, and wage rate. This fitted equation, evaluated at the mean of the data is equated to the net marginal benefits of seat belt use, up to a factor or proportionality, equal to the variance of the dependent variable. Assuming zero time and disutility costs of operation, the implied value of life is solved from this equation. His estimates of the average value of life, based on a non-random sample of about 5,500 households in A Panel Study of Income Dynamics, 1968-1974 is $370,000 ($ 1978) . However, Blomquist relies heavily of the estimated wage coefficient in the profit equation to estimate the variance of the dependent variable. To the extent that the wage rate does not accurately reflect value of life, these estimates will be biased.

Dardis estimates willingness to pay for risk reduction by examining data on consumers' voluntary purchase of smoke detectors and their expected reduction in the incidence of death by fire. He estimates the annualized cost of smoke detectors per household based on a catalog purchase price, life expectancy of ten years, an average of 1.5 smoke detectors per household, and discount rates of 5% and 10%. Then under the assumption that 13% of households in 1976 were equipped with detectors, that only 80% of these were functional, and that these functional detectors provided only 45% protection, the total deaths in the absence of functional detectors was estimated at 6,492. Savings of life from the provision of smoke detectors in each household was then estimated at 2,337 (equal to .8 x .45 x 6,492) for a probability of reduction in death of 3.16×10^{-5} for all households. Combining this probability with the annualized cost of smoke detectors yielded estimates of the value of life to purchasing households ranging from $293,000 to $341,000 ($ 1978). The estimated value of life to the entire population was considerably less - ranging from $157,000 to $175,000 ($ 1978).

Although the behavior of the general population is observed in these two studies of consumer safety devices, there are many important shortcomings to their work. The first two caveats associated with the wage rate willingness to pay studies also apply to the studies by Blomquist and Dardis. In

addition, the most serious problem with Dardis' approach is that the total value of consumer willingness to pay cannot be accurately estimated using the selling price of the safety device. Clearly, many consumers with higher subjective probabilities of risk would pay far more than the modest price of the detector, whose production costs are substantially lowered by scale economies. However, the empirical importance of this bias is not clear. In light of these shortcomings, we suggest the following theoretical structure for hypothesis testing in valuing health effects.

The problem of valuing health effects is the discovery of the rates at which individuals are willing to substitute air pollution-induced changes in health status for money or its equivalent. The conceptual framework employed in the great bulk of the work on the demand for health is the household production model, particularly its human capital versions [Grossman (1972), Crocker-Schulze, et al. (1979, pp. 137-149)]. In this framework, the individual or family unit is viewed as a firm attempting to maximize utility subject to constraints on the household budget and the production of goods and services which yield utility. Market goods and services are purchased and combined with the time of various family members in production. Household members are therefore implicit demanders of their own time resources as well as of the factors, including health status, that influence what they are able to do with these time resources. The framework is useful for studying the value of air pollution-induced health effects because: (1) it assesses individual well- being by "full income"--the value of all the individual's time, including time passed in productive nonmarket activities such as raising children--and not merely by his money income; and (2) it provides a means of introducing a priori information on behavior of organ systems into a health production function.

Within the household production framework, changes in behavior due to a change in air pollution-induced health status flow from three major sources. First, a change in health status can change the income and wealth positions of some individuals, thus changing the amount and possibly the mix of "commodities" these individuals consume. Second, changes in health status may influence the type of income sought by the individual. Individuals can be expected to shift their efforts and investment patterns toward obtaining those types of income that yield the highest net return for expended time and money. Alternatively, because of increases in the difficulty of internal financing, reductions in self-investment, job search schooling, on-the-job-training, and migration may occur. Finally, various income support programs as well as the individual's social reputation are contingent upon others' perceptions of one's health status. Therefore, to the extent possible, individuals will tend to tailor their self-reported health status to increase their chances of being categorized in a manner offering them the most advantageous time and money

terms.

Thus, changes in wage rates and income will reflect, to a degree, changes in health status. Wages, which are the most important source of income for most households, are fairly accurately reported in most data sets. However, this by no means implies that they are free of measurement error and other problems. There are" at least three major difficulties with most wage data.

First, the individual's behavior is based upon his marginal, not his average, wage rate. The marginal wage rate is net of taxes and it must be adjusted for fringe benefits and for the cost-of-living. Since marginal and average rates obviously differ for all persons subject to progressive income taxes, failure to take account of these taxes will bias toward zero the estimated coefficient relating hours worked to wages.

Second, the wage rate used for estimation should distinguish between the permanent and the transitory components of wages [J.P. Smith (1977)]. The observed wage rate may be systematically related to the wages the individual expects to receive in the future. Ignoring anticipations regarding wage profiles over the life cycle can lead to seriously biased results. For example, if people who currently receive relatively high wages anticipate more steeply sloped wage profiles than do low wage people, the effect of current wage on labor supply is likely to be underestimated. To help control for the effect of differences in permanent and transitory wages, J.P. Smith suggests estimating expressions using cross-sectional data on narrowly defined age groups.

Third, data must be provided that allows the imputation of wage rates for nonworkers, many of whom adopt this status because of health problems. One solution is to impute a potential wage rate for nonworkers on the basis of the wages observed for healthy persons of otherwise similar characteristics for whom wage data is available. Gronau (1974) shows, however, that this procedure will overstate wage rates for individuals belonging to groups with low labor market participation rates.

Changes in income can occur for reasons other than changes in the wage rate. In particular, it is necessary to know the individual's and the household's nonemployment income flows. For most households, the primary sources of nonemployment income are the home and the automobile. Ignoring the nonmonetized returns from these assets can seriously bias estimated relations between changes in income levels and changes in behavior. J.P. Smith (1977) suggests that the problem of imputing values to nonmonetized assets can be avoided if subsamples are defined to include individuals who are at the same point in their life cycles and have had similar wage paths and other factors

that may influence their time allocations over the life cycle.

Another important determinant of individual income is the amount of labor the individual supplies. Because of disequilibria in labor markets, the actual hours of employment for some persons may differ substantially from the number of hours they wish to work at the wage rate they receive [Ashenfelter (1977)]. When assessing the value of air pollution-induced changes in health status, we wish to know the changes in actual hours worked.

All of the above wage and labor supply responses may differ among various types of people; that is, the characteristics defining types of people may interact with the explanatory variables of the expressions to be estimated. When these characteristics are exogenous, and when the existence but not the form of the interaction is known, the sample must be stratified so that separate estimates can be made for each type. Failure to do so can lead to seriously misleading estimates. Crocker and Horst (forthcoming) have shown, for example, that reductions of the earnings of workers in the same occupation exposed to near-identical ambient concentrations in Los Angeles vary between zero and nine percent. Pooling these workers would have imposed statistically unacceptable restrictions. In light of the preceding discussion of the optimal use of prior information we draw the following conclusion: in the absence of prior information and hypothesis testing, the "ideal" data set cannot be specified. One can only say that data on all imaginable factors that affect health status, will not be ideal since it will produce intolerable risk. To minimize risk, we must introduce priors from accumulated statistical evidence to structure testable hypotheses about functional form; dimensionality of the parameter space, the model, the values of the design matrix under experimental control, and the density of the dependent variable, and we must employ optimal test procedures otherwise, there is no optimal way to judge the value of a data set. A good approximate specification of what would be ideal must therefore wait upon the results of explorations of what is gained by imposing more structure on existing data sets. For example, the introduction into the model structure of expressions for metabolic processes and organ system functions can provide identifying restrictions for the parameters of the self-reported disability, even though such data are scarce. Of course, the most complete identifying restrictions would be obtained if direct observations were available on these processes and functions. A data set having these observations could then be used to assess the gains from including expressions for these undeserveable processes and functions in the model structure relative to the gains from having direct observations on them. Given the likely expense of collecting accurate data on organ system functions, for example, a prior assessment of the size of these gains seems a worthwhile investment.

steps in the causal chain in Figure 1 as the HES data set. This is the Health Examination Survey (HES) data set collected from late 1959 through 1962 for a nationwide sample of 7710 adult, civilian, noninstitutionalized individuals [National Center for Health Statistics (1965)]. Given the early date of the HES data set and the broadness of its locational information (counties or sets of counties), more measurement error than usual would be introduced when the set was matched with air pollution information. However, as Leaderer (1979) has suggested, visibility information from airports might serve as a very adequate proxy for fine particles which are suspected as the major source of health impairment from air pollution.

CONCLUSIONS AND RECOMMENDATIONS

Neither epidemiologists nor economists are yet able to provide estimates of the health consequences of air pollution with sufficiently reliable hypotheses to carry out a defensible cost-benefits analysis. The range of uncertainty is unacceptably large. A traditional response to unacceptably large ranges of uncertainty is a plea for undertaking a fresh data collection effort. To say that one wants all "feasible" information on individuals' genetic and social endowments, metabolic processes, organ system functions, past and present life-style habits, risk exposures other than air pollution, attitudinal variables related to stress, indoor and outdoor air pollution exposures, family characteristics and employment opportunities, as well as a history of time and budget allocations is to say little. Minimization of estimator risk requires physiological and economic models to specify testable hypotheses and hence to guide the data specification. A great deal of relevant economic information will have been made available when the measures of labor supply, wages, and income described in the previous sections are generated. Smoking habit information, diet, and occupational exposures appear to be necessary. Beyond this, data sets must be collected and explored with the explicit objective of minimizing estimator form, model selection, experimental control of the design matrix, and choice of density function for the dependent variable. This will require that more attention be devoted to the role played by organ system functions using data disaggregated to the individual level. Expressions which purport to explain these functions, along with expressions which explain time and budget allocations, will most likely become the major sources of a *priori* information that can be used to bound the investigation. Thus, the epidemiologist is at the difficult position where more testable hypotheses appear to be as important as more data.