Volume V

AIR POLLUTION AND DISEASE: AN EVALUATION OF THE NAS TWINS

METHODS DEVELOPMENT IN MEASURING BENEFITS OF ENVIRONMENTAL IMPROVEMENTS

Volume V

AIR POLLUTION AND DISEASE: AN EVALUATION OF THE NAS TWINS

by

Ralph C. d'Arge
Susan B. Kask
James C. Case
University of Wyoming
Laramie, Wyoming 82071

Shaul Ben-David
University of New Mexico
Albuquerque, New Mexico 87131

Larry S. Eubanks
University of Oklahoma
Norman, Oklahoma 73019

Curtis L. Anderson
University of Minnesota
Duluth, Minnesota 55812

USEPA Grant #CR808-893-01

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

DISCLAIMER

This report has been reviewed by the Office of Policy Analysis, Office of Policy, Planning and Evaluation, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

CONTENTS

ABSTRACT

This analysis focuses on the relationship of disease symptoms to air pollution and other health effecting variables. Sulfur dioxide was found to contribute to the incidence of chest pain while increasing levels of total suspended particulates were found to increase reported incidence of severe chest pain, coronary heart attack, cough, and shortness of breath. Statistical measures used in this analysis were probit and ordinary least squares. Total estimates of the savings in health care costs due to a 30 percent reduction in ambient concentrations of TSP and SO_2 are derived and found to have substantial magnitudes.

FIGURES

TABLES

ACKNOWLEDGMENTS

This analysis was conducted with the partial financial assistance of the U.S. Environmental Protection Agency. The authors wish to thank Dr. Alan Carlin of EPA for his efforts to aid us in the performance of this study, and also his help in obtaining relevant data sets from the National Academy of Sciences. We wish to also thank Dr. Zdenek Hrubec of the NAS for his help in getting access to and using the Twins data set. Drs. Thomas Crocker, Shelby Gerking, and William Schulze have made valuable comments on the methods and approaches taken in this study. Also, Dr. Allen V. Kneese of Resources for the Future has made several incisive comments that we have utilized. The authors take full responsibility for all errors, omissions, and incorrect assumptions taken in this effort.

SECTION 1

INTRODUCTION

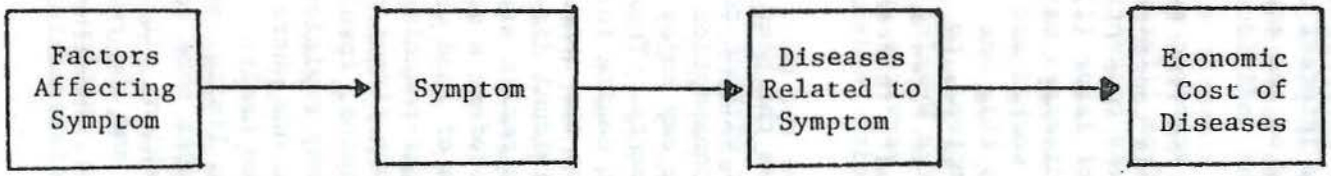
Human disease is caused by a mosaic of events, exposures, aging, psychoses, genetic background, and the environment in which the individual resides. Air pollution is but one of the many factors potentially influencing morbidity and mortality rates of the population. The central question arises as to 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 indiviudal 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 occurence 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 particulates 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.



Diet

Age

Asthma

Alcohol Consumption

Income

Cigarette Consumption

Ambient Air Pollution

Cough

Shorthess of Breath

Chest Pain

Severe Chest Pain

Coronary Heart Attack

Chronic Bronchitis

Emphysema

Bronchiectasis

Chronic Interstitial Pneumonia

Ischemic Heart Disease

Congenital Heart Disease

Cardiomyopathy

Cardiac Failure

Hospital Expenditures

Loss of Earnings

Loss of Earnings Due to Death

Physicians Services

Figure 1. Major Relationships Examined and Statistically Estimated for the NAS Twins

SECTION II

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 (1). 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 disutility. 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:

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

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,

$$U_Q,\ U_H,\ U_X > 0;\ U_{QQ},\ U_{HH},\ U_{XX} < 0\ . \qquad (2)$$

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,

$$U_C,\ U_B,\ U_E > 0;\ U_{CC},\ U_{BB},\ U_{EE} < 0\ . \qquad (3)$$

If they are viewed as "bads" then,

$$U_C,\ U_B,\ U_E < 0;\ U_{CC},\ U_{BB},\ U_{EE} > 0 \qquad (4)$$

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,

$$O2 = f(K, A) \tag{5}$$

where:

O2 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

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

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 both 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 of a given 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.

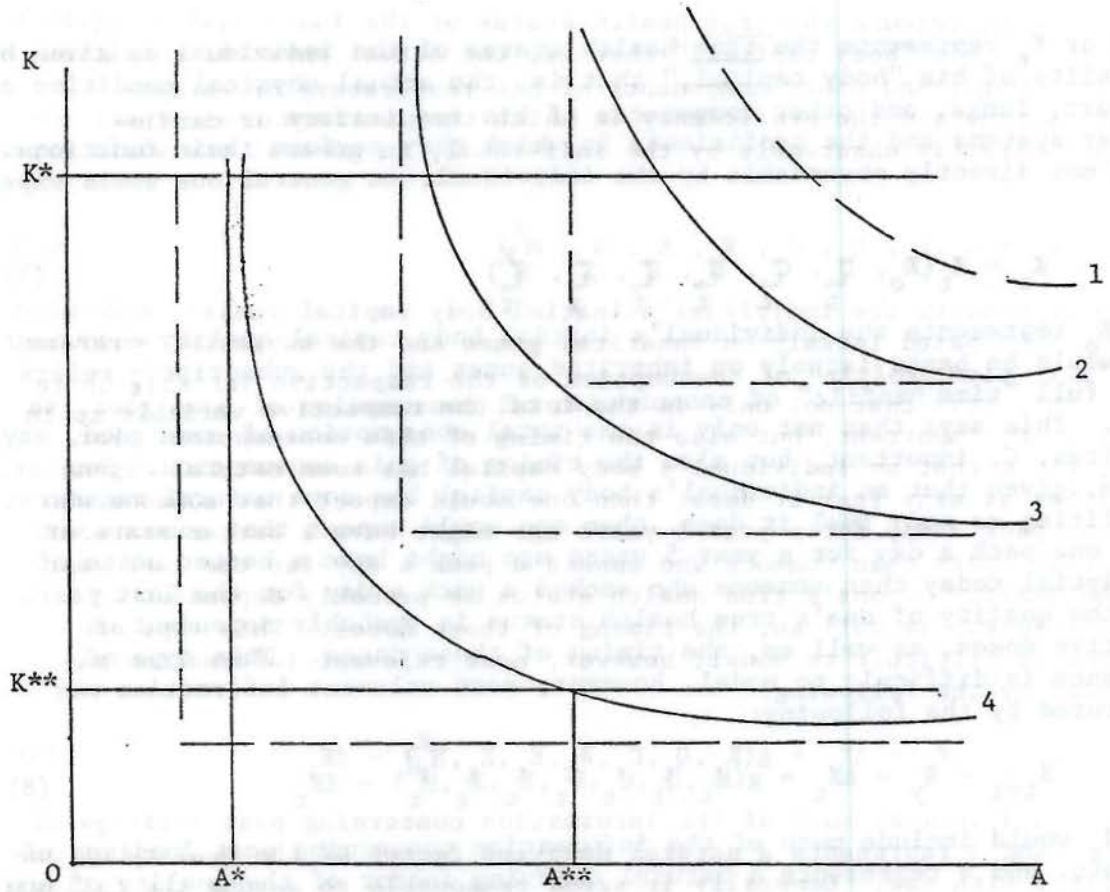


Figure 2. Conceptual Tradeoff Between Body Capital and Respiration

$^{1}O_2$ 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 <u>true</u> 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_o represents the individual's initial body capital quality endowment which would be based largely on inherited genes and the subscript $\underset{\rightarrow}{t}$ 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:

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

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 unuseable). 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 unuseable). 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

$$S_t = S_t\ (K_t,\ Q_t,\ M_t^S) \tag{10}$$

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

$$S_K,\ S_{Q_t},\ S_{M^S} > 0. \tag{11}$$

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

$$H_t = H_t\ (S_t,\ M^I_t) \tag{12}$$

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:

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

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

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

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

$$K(0) = K_o$$

where $\bar{Y}$ is the individual's income constraint and $P_{\cdot\cdot}$ 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 $U_C \gtrless 0$?) 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 occurences 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 $u(\cdot)$ 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_X 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:

$$\left.\begin{aligned} u_X - \lambda P_X &\leq 0 \\ u_S + \delta h_S &\leq 0 \\ -\lambda P_M - \delta &\leq 0 \end{aligned}\right\} u_S + \lambda h_S \leq 0 \tag{17}$$

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 <u>perceived</u>. A derived demand relationship for M can be developed from the presence of symptoms as follows:

$$M = f(P_X, P_M, S) \tag{18}$$

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 P_M. As an alternative, average U.S. medical expenditures for each type of illness were used to estimate a <u>minimum</u> 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.

SECTION III

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 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 3. NAS Twins (Q1) 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 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 not 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 particulates 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 particulates 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.

SECTION IV

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 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	Intoxication	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 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	Fiboflavin	mg per year, in 1967

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

TABLE 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
USFT	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 4. CORRELATION MATRIX

	RIBF	PROT	FATS	VITA	NIAC	THIN	CALC	IRON	NTRS	USFT	FIBR	SUGR	STFT	ZYGT	SO_2M	TSPH	EARN	BRIW	WT25	WGHT	HGHT	LIQR	CIGM	INTX	DRNK	SMKN	DIET	RUNF	ASTH	CORN	SVCP	COCH	SHAR	CHPH	YWHO
YWHO	-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	0.01	-0.01	0.01	0.00	-0.01	-0.01	0.00	1.00
CHPH	-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	0.12	0.22	0.19	1.00	
SHAR	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	0.05	0.08	0.07	-0.02	0.05	0.02	0.00	0.06	0.19	0.18	0.19	1.00		
COCH	-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	-0.00	0.02	-0.04	-0.01	-0.03	-0.04	-0.03	0.15	0.27	0.07	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	0.00	0.01	-0.02	-0.01	-0.03	-0.03	-0.02	0.00	0.05	0.02	-0.00	0.02	0.05	-0.01	0.00	0.35	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	0.03	1.00					
ASTH	0.01	0.02	0.02	0.01	0.02	0.02	0.01	0.02	0.03	0.01	0.00	-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						
RUNF	-0.02	-0.02	-0.01	-0.02	-0.03	-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	-0.00	-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										
EDUC	-0.02	-0.01	-0.01	-0.02	-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	0.12	0.14	1.00											
CIGM	-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	0.23	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	0.03	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	0.03	1.01	0.01	-0.01	0.01	0.06	-0.01	0.01	0.05	0.05	0.51	0.53	1.00														
WGHT	0.01	0.01	0.01	0.01	-0.01	0.01	-0.00	0.01	0.03	0.01	-0.02	-0.03	0.01	0.03	-0.01	0.01	0.04	0.01	0.73	1.00															
WT25	-0.02	-0.01	-0.00	0.00	-0.03	-0.01	-0.02	-0.01	0.02	-0.02	-0.02	-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	0.01	-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																		
TSPH	-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																					
STFT	0.90	0.90	0.93	0.64	0.76	0.59	0.80	0.86	0.21	0.97	0.20	0.71	1.00																						
SUGR	0.68	0.50	0.62	0.40	0.43	0.43	0.34	0.49	0.15	0.75	0.40	1.00																							
FIBR	0.28	0.29	0.21	0.19	0.39	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																									
[illegible]	0.33	0.39	0.42	0.14	0.41	0.76	0.21	0.35	1.00																										
IRON	0.92	0.98	0.89	0.79	0.87	0.79	0.80	1.00																											
CALC	0.87	0.79	0.84	0.84	0.56	0.53	1.00																												
THIN	0.70	0.79	0.74	0.46	0.81	1.00																													
NIAC	0.70	0.89	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 SO_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 SO_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 SO_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 SO_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 SO_2 or TSP.

TABLE 5. ALTERNATIVE ORDINARY LEAST SQUARES REGRESSIONS WITH 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
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)	-	-	-	-

Dependent Variable and Regression #	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

TABLE 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 II	.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	NIAC	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

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 particulates 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 7. ALTERNATIVE ORDINARY LEAST SQUARES WITH THE INCIDENCE OF CORONARY HEART ATTACK 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
CORN I	.1686 (7.248)	.0183 (.962)	-.0427 (-1.617)	.00002 (1.287)	-.0156 (-4.758)	-.2740 (-3.316)	.0002 (2.763)	-.0001 (-1.267)	-	-	-	-
CORN II	.1640 (7.017)	.0192 (1.009)	-.0450 (-1.701)	.00002 (1.006)	-.0155 (-4.722)	-.0028 (-3.359)	.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 (6.550)	.0250 (1.309)	-.0555 (-2.091)	.00002 (1.031)	-.0158 (-4.826)	-.0028 (-3.423)	.0002 (2.729)	-.0001 (1.335)	-	-	-	-

	NTRS	IRON	CALC	THMN	NIAC	VITA	FATS	PROT	RIBF	CONSTANT	R^2	SSR DF
CORN I	-	-	-	-	-	-	-	-	-	.4529 (5.701)	.0120	5737 7899
CORN II	-	-	-	-	-	-	-	-	-	.5364 (6.036)	.0126	5734 7898
CORN III	-	-	-	-	-	-	.0103 (1.281)	-.0227 (-2.845)	-	.6911 (6.425)	.0143	5724 7893
CORN IV	.5283 (1.705)	.2718 (1.548)	1.2481 (.734)	-.4602 (-2.073)	-.1206 (-1.268)	-.4820 (-1.420)	-	-	-.5041 (-1.360)	.5089 (1.483)	.0149	5720

TABLE 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_2M	STFT	SUGR	FIBR	USFT
COGH I	-.0166 (-1.873)	.0952 (13.873)	.0097 (.966)	.99996 (10.366)	-.0019 (-1.537)	-.0010 (-3.306)	.00006 (2.229)	-.00004 (-.896)	-	-	-	-
COGH II	-.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 III	-.0158 (-1.771)	.0950 (13.021)	.0121 (1.201)	.00006 (10.348)	-.0018 (-1.438)	-.0010 (-3.096)	.00006 (2.341)	-.00004 (-.865)	-.0119 (-1.739)	.0013 (2.108)	-.0944 (-5.513)	.0051 (.758)
COGH IV	-.0102 (-1.132)	.0931 (12.792)	.0103 (1.024)	.00006 (10.308)	-.0018 (-1.474)	-.0009 (-3.012)	.00006 (2.350)	-.00004 (-.963)	-	-	-	-

	NTRS	IRON	CALC	THMN	NIAC	VITA	FATS	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 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_2N	STFT	SUGR	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 III	.0256 (2.242)	.0350 (3.761)	-.0400 (-3.099)	.00004 (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)	-.0393 (-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 III	-	-	-	-	-	-	-.0123 (-3.140)	.0106 (2.714)	-	.0669 (1.276)	.0103	1360 7893
SHBR IV	-.4710 (-3.117)	-.0670 (-.782)	-1.0787 (-1.301)	.1877 (1.734)	.0283 (.609)	.0931 (.562)	-	-	.2437 (1.348)	.2258 (1.350)	.0102	1361 7892

TABLE 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 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 particulates 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 SO_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 particulates 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. SO_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 SO_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.

SECTION V

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 particulates, 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 particulates 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 particulates and sulfur dioxides as pollution variabiles. Using both linear and non-linear models, they found that SO_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:

$$P(D) = P(S_y) \cdot P(D/S_y) \tag{19}$$

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:

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

where $\tilde{P}_o$ 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:

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

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

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

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:

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

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

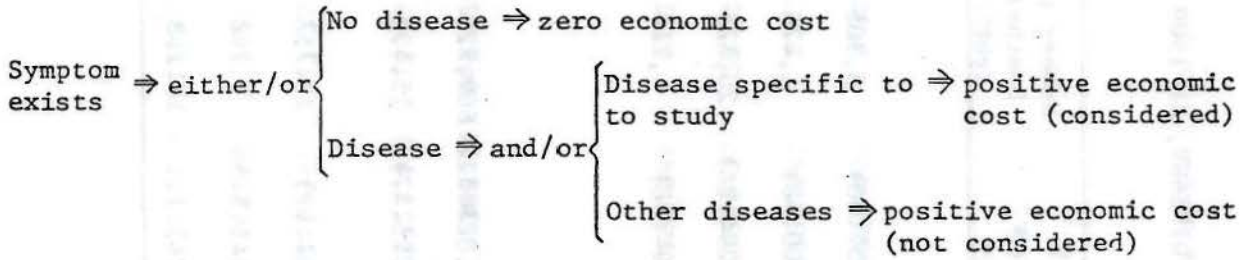
$$\Delta E(C_D/\Delta\tilde{P}) = \Delta P(D/\Delta\tilde{P}) \cdot C_D. \tag{24}$$

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

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

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 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 Intestitial 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
Arrythmias	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.

[c]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.

[e]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 $\cdot$ 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 undoubtably 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:

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

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 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 Intestitial 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_o and a change in the pollution level occurs bringing society to a level of P_1^o the benefits received are B_1. Now if the initial level is $\tilde{P}_o$ 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 catagorization 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 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 Intestitial 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		
	Arrythmias	.00009		
	Coronary Heart Attack	.00017		
Chpn	Chronic Bronchitis	.00076		.00693
	Bronchiecstasis	.00004		
	Emphysema	.00083		
	Ischemic Heart Disease	.00486		
	Cardiomyopathy	.00022		
	Arrythmias	.00022		

TABLE 14 (continued)

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

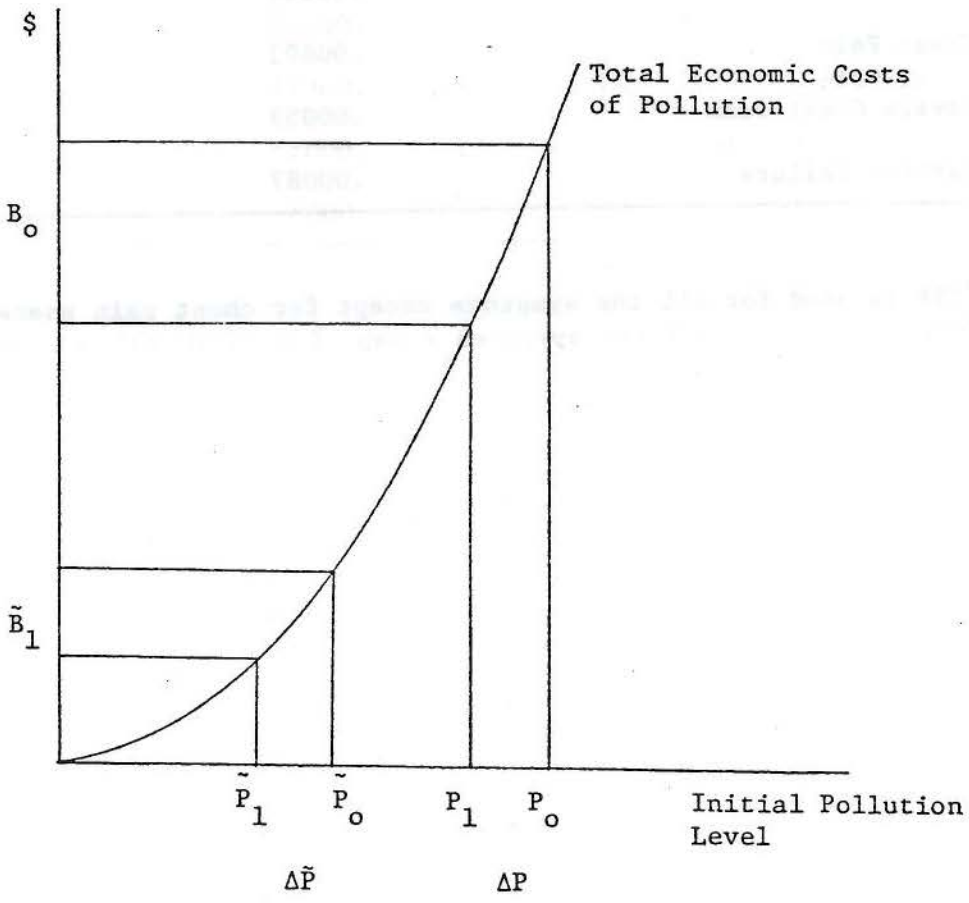
[a] values are reported only if the regression coefficient has a positive sign

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

TABLE 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 SO_2 is used.



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

Figure 5. Measuring benefits from pollution reduction assuming increasing costs of pollution

TABLE 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 particulates. 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 particulates. Urban benefits of reduced mortality due to a 60 percent reduction in the level of total suspended particulates 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.

SECTION VI

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 particulates, 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 particulates 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 SO_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

$$1.5 \text{ servings/month} \cdot 7 \text{ gr/serving} \cdot 12 \text{ months/year} = 126 \text{ gr/year}.$$

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 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	0	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	511.0	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	7.30	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	126	0	0	.12	3.0	.66	24	2.4
Eggs[3]												
(64) 1	4380	4380	2190	1460	-	-	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	0	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	133.2	4.8	-	.30	12	.78	60	4.8

TABLE 15 (continued)

Fruit[3] (67) 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 be identified by response. Var. 53-59 do not have response included in the calculation and therefore the coefficient must be multiplied by the response.

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

Response	Assumption
0 never	0 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 not available but the nutrient is suspected to be present
- : only a trace has been detected
0 : the nutrient is not present and is not 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 to Good Eating, 1980

TABLE 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 19. LEVELS OF NUTRIENTS AND NITROSAMINES PER SERVING BY TYPE OF FOOD

Nutrients / Type of Food	Serving	Protein (gm)	Fats (gm)	Fatty Acids uns (gm)	Fatty Acids sat (gm)	Carbohydrates sugar (gm)	Carbohydrates fiber (gm)	Vit. A (1u)	Riboflavin (mg)	Niacin (mg)	Thiamin (gm)	Calcium (mg)	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
Coffee[1] (56)	1 cup*	.3	.1	na	na	.8	0	0	.01	.9	.01	4.6	.23	0
Coffee w/ tsp. sugar (57)	1 cup w/1 tsp. sugar	.3	.1	na	na	11.8	0	0	.01	.9	.01	4.6	.23	0
Tea[1] (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 cup 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	2.2	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 daily recommended servings for these variables. We assumed the average serving of pastery as 1 and an average serving of candy as a candy bar

References: 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 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.

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. Leavell, 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-600/5-79-001a, 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. Leavell, 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.

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

U.S. 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.