AN ABSTRACT OF THE THESIS OF

JOHN AUGUST JAKSCH for the DOCTOR OF PHILOSOPHY
(Name of student) (Degree)

in Agricultural Economics presented on October 5, 1972
(Major) (Date)

Title: OUTPATIENT MEDICAL COSTS RELATED TO AIR POLLUTION IN THE PORTLAND-VANCOUVER AREA

Abstract approved: ______________________
Dr. Herbert H. Stoevener

The main supposition of this research was that air pollution can aggravate a state of health resulting in increased consumption of outpatient medical services, and in the number of contacts with the medical system, for certain respiratory, cardiovascular, and other diseases aggravated by air pollution. The main objective of this study was to quantify such effects, if any, and express them in monetary terms. The study examines the effect of day to day changes in air quality on the consumption of outpatient medical services. It was felt that, in order to isolate the effect of air pollution on health, account must be taken within the explanatory models of other variables which could influence the consumption of medical services. These included socioeconomic-demographic and meteorological variables.

The study period was for 1969 and 1970, and centered in the Portland, Oregon Metropolitan Area. Medical and

socioeconomic-demographic data were obtained from the Oregon Region, Kaiser Foundation Health Plan, a prepaid group medical plan with 150,000 members, through Kaiser Health Services Research Center. The medical and socioeconomic-demographic data were obtained from a five percent, ongoing, random sample of the Kaiser membership. The air quality data used in the study were suspended particulate matter, and were obtained from several air pollution control authorities located in Portland, Oregon and Vancouver, Washington. The meteorological data used in the research came from the National Weather Service.

It is argued that in the consumption of medical services for cardiovascular-respiratory problems, the price of medical services and the patient's income may be discounted as variables affecting the demand for medical services regardless of the medical care source. The argument is supported by empirical evidence. The bases of support for the argument, for purposes of this study, are as follows: the nature of the disease being examined (without medical treatment certain cardiovascular-respiratory diseases will terminate in death); a minimum amount of medical services required to maintain the patient's health state who is suffering from the diseases; with respect to price, the inability of the patient to effectively evaluate alternative medical procedures at different prices. Hence, it is concluded, the use of a prepaid system such as Kaiser's can be used to estimate the medical costs of certain diseases associated with air pollution, and the estimates derived will be representative of other medical systems.

All the outpatient medical services used to treat a single disease incident per contact with the medical system were expressed in California Relative Value Units. The units were transformed to dollar values such that the dollar value would approximate the quantity of medical services consumed per contact.

The regression models (Models (8.1) and (8.2)) were formulated and were tested with the outpatient medical services consumed in treating respiratory and circulatory-respiratory diseases. The dependent variable in Model (8.1) specified the consumption of outpatient medical services per contact with the Kaiser system. The dependent variable in Model (8.2) expressed the medical costs on a per capita basis per census tract. Both models contained the following variables for each patient: age, sex, marital status, number of people per household, household income, race, drinking characteristics, cigarette smoking characteristics, occupational exposure to job related pollutants, ambient air quality, and meteorological conditions (expressed as a Temperature Humidity Index). A measure of physical fitness was included in Model (8.1), but not in Model (8.2). The independent variables in Model (8.2) were averaged over census tracts. The air pollution and meteorological variables in both models were lagged up to three days to reflect the fact there may be a delay between exposure and disease onset. This resulted in four sub-models for the respiratory and circulatory-respiratory disease categories in (8.1). Model (8.2) was tested only against respiratory diseases, and lagged one day. The statistical results are summarized as follows.

Model 8.1 - Respiratory Diseases

The following regression coefficients were statistically significant at the 20 percent level or higher in all four sub-models, except where noted: sex, negatively related to the dependent variable; drinking habits, positively related to the dependent variable; ambient air pollution, significant in all models, except Sub-Model 4 (three day lag), and all coefficients were positively related to the dependent variable; meteorological conditions, positively related to the dependent variable.

The coefficients of multiple determination (R^2) for Sub-Model 1 through 4 were 0.022, 0.024, 0.022, and 0.022, respectively. The overall F tests for regression were statistically significant at the one percent level for all models. The air pollution variable was statistically significant from zero at the two percent level when lagged by one day (Sub-Model 2). Increasing air pollution by 20 micrograms, from 60 to 80 micrograms per cubic meter, resulted in an estimated 3.5 cent increase in outpatient medical costs per contact with the medical system. The low R^2's were attributed to the type of medical services that were being analyzed and the specification of the model.

Model 8.1 - Circulatory-Respiratory Diseases

The following coefficients were statistically significant at the 20 percent level in all four sub-models, except where noted: age, significant only in Sub-Model 1 - unlagged - and with a negative instead of the expected positive sign; physical fitness, negatively related to the dependent variables; drinking habits, positively related to the dependent variable; occupational exposure index, positively related to the dependent variable; meteorological conditions, positively related to the dependent variable.

The R^2's for each of the Models 1-4 were 0.0080, 0.0083, 0.0076, and 0.0080, respectively. The overall F tests for regression on all four models were statistically significant at the five percent level.

Model 8.2 - Respiratory Diseases: One Day Lag

The following variables were statistically significant at the 20 percent level or higher: sex, positively related to the dependent variable; household income, positively related to the dependent variable instead of the expected negative relationship; drinking habits, positively related to the dependent variable; smoking index, positively related to the dependent variable; meteorological conditions, positively related to the dependent variable. The overall F test for regression was barely statistically significant at the five percent level. Primarily because the dependent variable did not exhibit much variation, the results of this model should be interpreted with caution.

The results from Model (8.1) infer that air pollution does have some effect on the consumption of outpatient medical services used to treat respiratory illnesses. The results would indicate that the procedures outlined in this study appear to hold some promise in quantifying such effects. However, it is quite evident that the results leave much to be desired, and a great deal of work needs to be done. Primarily this would include retesting the models, respecifying some of the variables within the models, and expanding the research to include the effects of air pollution on the consumption of inpatient medical services.

Outpatient Medical Costs Related to Air Pollution In the Portland-Vancouver Area

by

John August Jaksch

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

June 1973

APPROVED:

Professor of Agricultural Economics
in charge of major

Head of Department of Agricultural Economics

Dean of Graduate School

Date thesis is presented October 5, 1972

Typed by Ilene Anderton for John August Jaksch

ACKNOWLEDGEMENTS

The author is indebted to many persons and organizations for advice and assistance in conducting this research. In particular, debts of gratitude are due to:

my wife, Carol, for continued encouragement, thoughtfulness, help, and sacrifice throughout the two years of this research.

Dr. Merwyn Greenlick, Director, Kaiser Health Services Research Center, for providing the medical and socioeconomic-demographic data used in this study, and to Dr. Greenlick's staff for the unlimited assistance they gave me in obtaining and preparing the data for analysis.

the staff at Columbia-Willamette Air Pollution Authority, Department of Environmental Quality, and Southwest Air Pollution Control Authority for providing the air pollution data used in this study, and for giving unlimited technical advice.

Mr. Stanley Holbrook, State of Oregon Climatologist, for providing the meteorological data used in this project.

Dr. Herbert H. Stoevener for his guidance, helpful criticisms, and valuable assistance to this research in his role as major professor.

Mr. James Sasser, my computer programmer, who often gave freely of his time, and worked well beyond the normal eight hour day to program the data for analysis.

Sheldon Wagner, M. D., who willingly served as the medical consultant for this study.

Dr. John A. Edwards who contributed many ideas incorporated within the theoretical framework of this research and who helped me personally in many other ways.

Gladys Grammer, Opal McKinney, Jeanne Garst, and Dolly Sawtelle for the several months they spent coding portions of the data used in this study.

the Environmental Protection Agency for research and stipend support to carry out this project (fellowship number 5 FO 3 AP 38677 and contract number 68-01-0423).

and the staff members, Dr. Richard W. Boubel, Dr. Roger G. Petersen, Dr. Joe B. Stevens, and Dr. Tjeerd van Andel for their advice and assistance as members of the Ph. D. committee.

TABLE OF CONTENTS

Chapter | Page

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDIX TABLES

OUTPATIENT MEDICAL COSTS RELATED TO AIR POLLUTION IN THE PORTLAND-VANCOUVER AREA

I. INTRODUCTION

The Problem

For informed and economically efficient decisions to be made in managing the environment, adequate information must be provided to the decision-maker. One of the main problems of air quality management is balancing the benefits against the costs that society derives from using the atmosphere as a dumping ground for its wastes. In attempting to move toward this balance, the decision-maker must have estimates of the costs and benefits associated with achieving certain environmental quality levels. Attempting to provide answers for these questions generally involves difficult issues in measuring and evaluating the diverse benefits and costs of pollution abatement. It is in the need for such analyses that the role of the economist within environmental decision-making has become increasingly important.

While dependable, systematic estimates of costs resulting from the effects of air pollution are still quite rare, progress is being made. Within the past decade several studies have been completed, estimating property and material costs of air pollution and the effects of pollution on property values (30, 41, 58). One of the more important costs yet to be measured, however, is the medical costs attributable to air pollution. There is considerable evidence in the medical literature to support the hypothesis that air pollution affects human health and, particularly, adversely affects certain diseases of the respiratory and cardiovascular system (24, 84, 85). Adverse health effects of air pollution result in economic losses of unknown magnitudes. These losses to society are derived from two main sources: (a) the medical costs incurred in treating the disease, and (b) the value of the ill person's foregone production while sick. It will be the purpose of this research to estimate some of the outpatient medical costs attributable to air pollution.

Prior studies which have demonstrated an effect of air pollution on health have had one or more deficiencies. The deficiencies have been the quality of the air pollution, meteorological, socioeconomic-demographic, or medical data used, or in some instances not used, by the studies. The data bases afforded to this research appeared to offer a unique opportunity to overcome some of the deficiencies inherent in the data used by prior studies.

More specifically, there were available in Portland, Oregon (the study area) good air pollution, meteorological, medical, and socioeconomic-demographic data. The air pollution data were available from several air pollution control agencies. The data were representative of the main pollution problem within the study area, which is suspended particulate matter. There were enough ambient air quality stations to obtain a fairly good estimate of pollution conditions in any part of the area of study. There were adequate meteorological data from several National Weather Service Stations located throughout the area. An excellent source of medical data were available from Kaiser Foundation Health Plan (a group, prepaid medical plan), which allowed the research effort to concentrate on the effects of air pollution on the quantity of outpatient medical services consumed. This is opposed to other research efforts which have examined only the effects of air pollution on the number of mortality or morbidity cases per disease category, the number of doctor office visits, or the number of hospital admissions. Along with the medical data there were detailed socioeconomic-demographic data maintained by Kaiser on a five percent, ongoing, random sample of its total membership.

These bases provided a unique opportunity to overcome some of the deficiencies which have characterized other studies attempting to isolate the effects of air pollution on health. One of the major efforts of this research undertaking was to bring these multiple sources together, such that they all could be brought to bear on isolating the effects of air pollution on the consumption of outpatient medical services.

The social benefits of quantifying the outpatient medical costs of air pollution would be many.

(1) Decisions in environmental quality management should find part of their basis in economic analysis, as well as the current approach of focusing primarily upon the physical-health effects of pollution. One reason, perhaps, for an emphasis on the physical-health effects of pollution is the dearth of economic information translating such effects into economic costs. This study is an attempt to partially fill the void.

(2) It is not known if the economic losses resulting from the <u>usual levels</u> of air pollution (day-to-day changes, for example) are of greater significance for air quality management than the losses resulting from less frequent but more severe air pollution episodes. The question which needs to be answered is whether air pollution, at prevailing concentration levels, presents enough (additional) stress on the human body to contribute to the incidence, prevalence, and severity of certain diseases (60, p. 362). In essence, is air pollution harmful to human health at ordinary levels of concentration in the urban air as well as at the high levels which occur during a pollution episode (26, p. 589). This study will attempt to estimate the effect of air pollution on the consumption of outpatient medical services for day to day changes in air quality.

(3) One much recognize that certain socioeconomic-demographic characteristics can also play a role in affecting expenditures for medical care. Knowing the contribution to outpatient medical costs of certain levels of air quality and socioeconomic-demographic conditions would be valuable in establishing defensible criteria against which to judge competing public expenditure proposals in the allocation of scarce resources. That is, perceptiveness in this area could provide more reliable information upon which to base policy formulation on the allocation of resources between air quality control and the modification of socioeconomic-demographic conditions affecting health.

(4) A metropolitan area will usually have varying levels of air quality. By observing changes over time in air quality within given areas of the city with changes in the consumption of outpatient medical services used to treat certain diseases aggravated by air pollution, the local pollution control authorities will be able to determine what effect, if any, their efforts to control local pollution sources have on the occurrences of these diseases.

Research Objectives

The objectives to be accomplished by this research are as follows:

(1) To identify certain cardiovascular, respiratory, and other diseases which can be aggrevated by air pollution.

(2) To estimate the effect of air pollution, if any, on the consumption of outpatient medical services used to treat the diseases.

(3) If there is an effect of air pollution on the consumption of outpatient medical services, to quantify this effect in monetary terms.

(4) To determine the effect, if any, air pollution has on the incidence rate of the diseases identified in (1) above.

An Overview of the Research

The study area of this research is the Portland Standard Metropolitan Statistical Area for the period 1969 to 1970. The data for this study came from several sources. Kaiser Health Services Research Center, associated with Kaiser Foundation Health Plan which is a group, prepaid medical plan with 150,000 members, provided the outpatient medical and socioeconomic-demographic data. The outpatient medical records were obtained from a five percent, random sample of the total Kaiser membership. The socioeconomic-demographic data came from a questionnaire administered by Kaiser to the sample members during 1970 and 1971.

The air pollution data used in this study came from the Department of Environmental Quality, the Columbia-Willamette Air

Pollution Authority, and the Southwest Air Pollution Control Authority. The Department of Environmental Quality and Columbia-Willamette are Oregon air pollution control agencies; Southwest Air Pollution Control Authority is a State of Washington agency. The measure of air quality used in the study was suspended particulate matter. The meteorological data were provided by the National Weather Service.

Chapter II of this study presents the economic framework, and the statistical models and hypotheses to be tested. The study area and Kaiser Health Plan are described in Chapter III. Chapter IV details the compiling of the medical data, and the construction of the dependent variables in the statistical models. The collection of the socioeconomic-demographic data and the formation of the socioeconomic variables included in the statistical models are discussed in Chapter V. Chapter VI describes the Admatch program, and how it was used as a basis for assigning air pollution values to patients. The collection of the air pollution and meteorological data, and the construction of the air pollution and meteorological variables from the data base are detailed in Chapter VII. Chapter VIII is a presentation of the data analysis. A summary of this research, conclusions, and suggestions for future research are discussed in Chapter IX.

II. THEORETICAL FRAMEWORK

The Economic Model

Introduction

This chapter begins by specifying some of the assumptions used to formulate the economic model upon which this study is based. A graphical presentation of the economic model is given. From the graphical specification two important implications are drawn which minimize the importance of price of medical services and income in attempting to explain the consumption of medical services for certain cardiovascular-respiratory diseases.

In order to isolate the effect of air pollution on health account must be taken of other variables which could influence the consumption of medical services. A general framework is specified which delineates some of these characteristics. From this general framework, the hypotheses and the statistical models to test them are presented.

Specification of Assumptions

It is assumed that there exists a representative consumer-patient who is suffering from, or has a tendency to become afflicted by, certain diseases associated with air pollution. The consumer, having or contracting the disease, is a free agent to enter the market for medical care. There are no physical constraints which would prohibit him from obtaining such care.

This study is concerned only with the short-term effects of air pollution on pre-existing (chronic) and acute cases of certain respiratory, cardiovascular, and other diseases. These diseases are thought by the medical profession to be aggravated by air pollution, or have been shown by previous studies to be associated with air pollution.

Diseases of the cardiovascular and respiratory system affect two essential life systems of the body. Many of these diseases, if left untreated, are fatal or inflict such pain, suffering, and mental anguish upon the victim as to result in a decrease in the enjoyment of life. It is assumed that any consumer presently having any of the diseases, or newly contracting one of them, would be aware of these consequences. Given the choice between the alternatives, it is assumed that the representative consumer will actively seek medical care to maintain or improve his state of health.[1/]

It is argued the consumer is unable to effectively evaluate

[1/] For this to hold it is necessary to assume that life is preferable to death, and that, if given a choice, one would choose to be free of pain and suffering. A second supposition is the consumer expects the consumption of medical services to alleviate his illness.

medical care, and has to rely primarily upon the physician's advice for the treatment required to remedy the illness. This follows for several reasons. Most consumers do not have the expertise to effectively evaluate the alternative bundles of medical services used to treat their disease, or to judge the quality of physician services provided to them. Even if the consumer's disease is re-occuring, it would still be difficult for the patient to judge the quality of treatment received. It is highly unlikely he would know the exact, optimal course of treatment for each incident. There are so many variables which affect diseases from one incident to the next that, in many instances, it is difficult even for the physician to determine what to do.

Each disease incident will have a different degree of acuteness or severity, requiring a different intensity of treatment. It is argued that there is a certain minimum amount of medical care, given the type of disease and its degree of severity, which will maintain or improve the consumer's state of health.[2/]

[2/] A consumer's state of health will change over time, and will be dependent upon many socioeconomic-demographic characteristics, as well as the disease contracted. The consumption of medical services may do no more than preserve a patient's health state, or retard the rate of change toward further deterioration (terminal cancer, for example, where treatment just prolongs the inevitable). In other instances, the consumption of medical services may cure the patient of the disease and improve his health state.

In examining the patient's state of health, it is necessary to assume that the physician will do all that is necessary to preserve

It is recognized that the exact intensity, kinds, and quality of medical services might vary depending upon the different sources providing the care, as between different physicians, clinics, and hospitals, for example. With respect to the Kaiser system, such differences should be minimized since almost all medical services are provided within the system.

There are several ways that the rendering and consumption of medical services can be examined. One method is to view and analyze medical care as consisting of distinct treatment components, each with its own price (for example, a number of hospital days at a given price per day). Another method, used in this study, is to view medical services as a bundle of care components, the bundle being used to treat a single disease incident. The entire bundle of treatments is purchased to treat the disease incident. The use of this concept will become clear in the next several sections.

Graphic Specification of the Economic Model

A consumer's expenditures on medical care evolve from his preference map for medical services and other goods. The shape

---

or improve his patient's health state. (This is subject to the patient's socioeconomic-demographic-disease states which affect the patient's state of health, and are beyond the control of the physician.) Hence, when discussing maintaining or improving a person's state of health, it is meant that the health state is improved to the best extent possible.

of his indifference map will be different, *ceteris paribus*, depending on whether he is ill or not. If he is ill, the shape of the indifference curve will be affected by the severity of the disease state, which can be influenced by the quality of air he is breathing.

In Figure 1, money is shown on the vertical axis and medical services are on the horizontal axis. The units of medical services are expressed as an index. The index denotes the physical quantities of all medical services consumed to treat an illness. A more severe disease state could require that more medical services, and perhaps more intense use of the services, would be needed to treat the illness as compared to what would be used to treat a less severe state of the same disease, *ceteris paribus*. Excluded from the definition are items which would be considered frills of medical care not needed for the treatment of the disease state.

Assume, at the beginning of time period t, the representative consumer is *not ill*. This is reflected by the shape of his preference function in Figure 1. In Figure 1, I', I'', and I''' are indifference curves between money and medical services. The usual assumptions about the indifference curves hold: utility is constant for movements along an indifference curve; a higher indifference curve is preferred to a lower indifference curve (I''' is preferred to I'' and I'; I'' is preferred to I'); the indifference curves do not intersect; the indifference curves are concave from above.

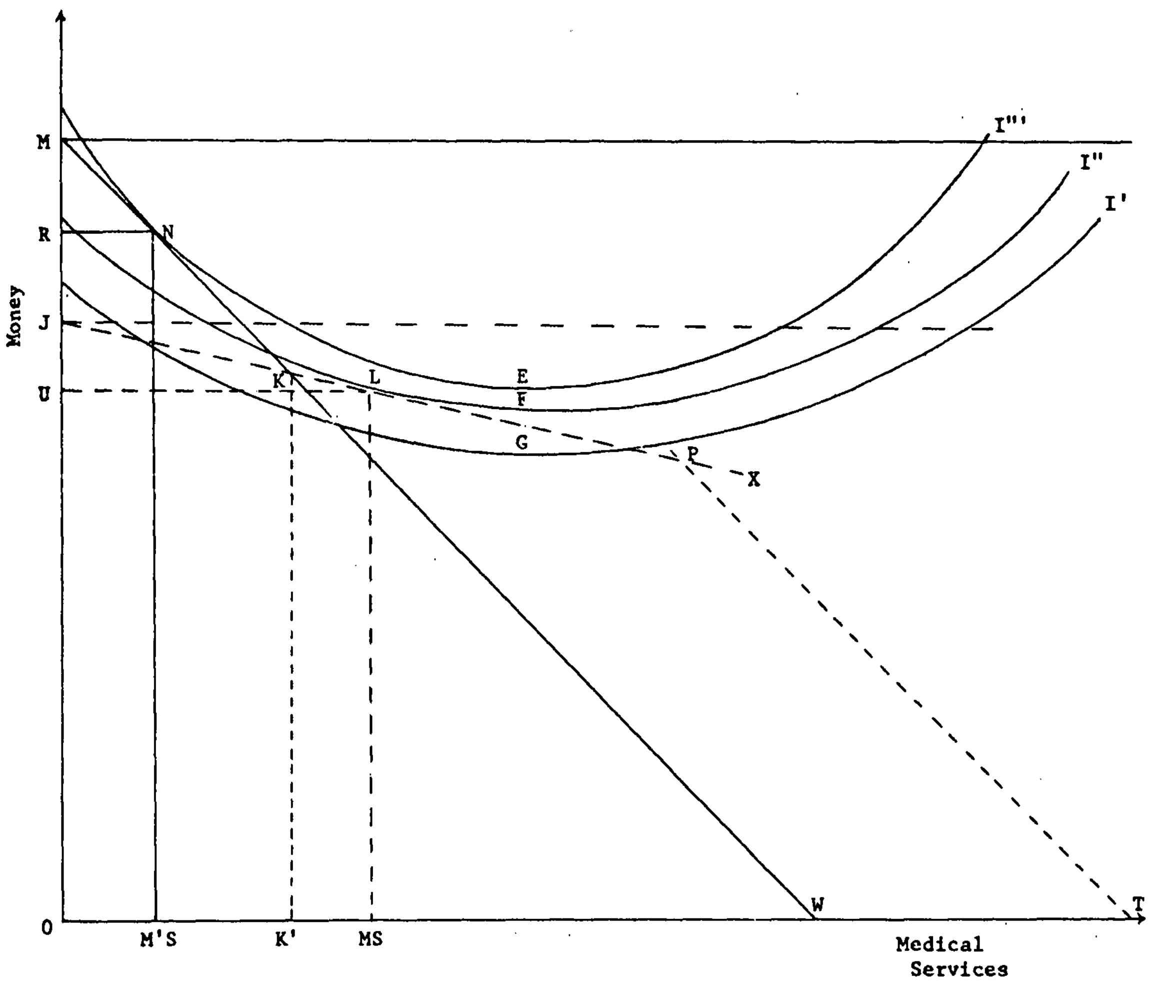

Figure 1. Preference map of a well consumer.

The slopes of the indifference curves play an important role in the analysis. While the consumer may not be ill at the beginning of this time period, he may still need medical treatment of a preventive nature (immunization shots and yearly physical checkups, for example), or anticipate the consumption of some medical services in case he does get ill. Hence, given his circumstances, the consumer has a rapidly declining marginal rate of substitution of medical services for money. His need for medical services is not a life-or-death situation; or a choice between continued physical pain and suffering and relief via medical treatment. In essence, the consumer's need for medical care is not urgent. It is quite possible that the consumer may become satiated with medical services, and the services may become a discommodity as the inconvenience of obtaining more services increases. In Figure 1 the points of satiation are illustrated by points E, F, and G on I''', I'', and I', respectively.

In fiscal year 1961, three-fourths of the population in the United States had some form of voluntary health insurance. The justification for having health insurance is the variability and unpredictability of disease incident. This can lead to broad fluctuations in individual and family expenditures on medical care over time, and to extreme differences among individuals and families at any given time (35, p. 31). A partial solution, of course, (in effect providing risk spreading) is health insurance. The consumer's decision to purchase insurance is his subjective evaluation of the quantity of medical services he _expects_ to use in time period t. This is based, in part, on the amount of medical services consumed in the past, on the cost of the insurance, and on the cost of medical care with and without insurance.

There is one thing to remember about medical insurance. Ex post, the consumer will always be worse off having bought the insurance, unless he becomes more severly ill than he had anticipated. However, he may feel better off, ex ante, having bought the insurance. It all hinges on the amount of medical services the consumer expects to consume in the time period for which he is considering an insurance contract, in relation to the actual amount of medical services required during that period. This is illustrated below.

OM in Figure 1 is defined as the consumer's permanent income. MJ is the amount of money the representative consumer would spend on a health insurance policy if he were to purchase it, leaving him the amount OJ to spend on all other goods and services. MW is the budget constraint reflecting the price of medical care (including any other private cost) to the consumer when he is not covered by insurance. Similarly, JX reflects the price for medical care when the consumer is covered by insurance.[3] Most insurance contracts have dollar or service limits within their provisions which, when exceeded by the insured, results in his having to pay the full cost of the excess services provided. In Figure 1, the limit of the policy is represented by point P. Hence, the consumer's budget constraint with insurance is JPT. Up to P, the price of medical care to the consumer is reflected by JP. When the limits of the policy are reached at P, the insured bears the full cost of medical care, represented by the price line PT which is parallel to MW.[4]

Given the conditions reflected in Figure 1, the crucial point in the decision to purchase medical insurance is point K. If the consumer feels that his consumption of medical services in t would be less than K', then he would be better off <u>not</u> to purchase insurance. Up to K', the total monetary outlay for the same quantity of medical services would be less without insurance than with insurance (premium cost MJ plus the cost of medical services with insurance, as reflected by the price line JP). Alternatively, if the consumer expects to purchase more medical services than that represented by K', then he can <u>expect</u> to be better off purchasing insurance.

---

[3] The slope of JX would shift with changes in the private and out-of-pocket cost of the insured.

[4] It is assumed, for purposes of this discussion, that the limits of the policy are not exceeded. Thus JP would be the relevant portion of the consumer's budget constraint.

Suppose the consumer decided not to purchase medical insurance. From this, one could infer he expects to be relatively healthy over time period t, and what medical services he anticipates consuming would be of a minimum or preventive nature best met out-of-pocket. A consumer would attain a position of utility maximization, which could shift with any change in his requirements for medical care as determined by his physician. The more insight a consumer has into his own medical needs during t, the more able he is to decide whether to purchase insurance. The point of utility maximization in Figure 1, for the consumer deciding not to purchase insurance, is at N on indifference curve I'''. The consumer would consume M'S of medical services and pay MR of his income for it.

If the consumer does purchase medical insurance, then one can infer he expects there might be a high enough probability of consuming enough medical services to make it pay; i. e., he expects to be better off having purchased the insurance contract. Again, a position of utility maximization could be tentative. Two reasons exist for this: (a) The consumer, at the beginning of t, is not ill but he is purchasing insurance against the possibility that he will become ill enough to consume an amount of medical services to the right of K'. The type of illness a person has, and relative demand for medical care, will vary for different disease categories, and hence be reflected in the consumer's preference structure. That is, the actual point of maximization may change because the consumer's indifference map changes when he becomes ill. (b) The quantity of medical services consumed by the patient, while ill, will be determined by the medical profession; i. e., the consumer, in most instances, does not know what diseases he may contract in t, or the amount of medical services needed to treat the disease. However, as in the non-insurance case, everyone has some subjective feelings as to the amount of medical services they might consume, on the average, in t. The point of utility maximization in Figure 1, for the representative consumer deciding to purchase insurance, is at L on indifference curve I''. The consumer would consume MS of medical services and pay JU, in addition to the premium cost MJ, for it.

Admittedly this framework is an oversimplification of what actually happens. It fails to address the question of the consumer's decision to buy insurance and his feelings about uncertainty. However, the argument is not central to the case being made by this study; i. e., air pollution can affect the consumption of medical services. The theoretical model presented here is a single period model. The model would be unduly complicated by addressing the question of uncertainty.

A prepaid medical plan is a form of medical insurance, where, as with other insurance plans, the mode of premium payment can vary the consumer's out-of-pocket costs for the premium between points M and J in Figure 1. The method of economic analysis would be the same in each case. From a welfare standpoint, ceteris peribus, the insured's position would change with each insurance plan and mode of payment.

Assume it is now sometime later in t, and the same representative consumer has become ill with a disease aggravated by air pollution. Reference is made to Figure 2, where MS_{a_1} is defined as the amount of medical services needed in attempts to preserve, improve, or prevent further deterioration of the consumer's state of health[5/] (given the kind of disease, its acuteness, and a level of air quality a_1). Except where it is noted, everything in Figure 2 is defined as in Figure 1. Any treatment to the left of MS_{a_1} represents less treatment than is needed to maintain the patient's state of health. Treatment in this area could result in deterioration of the person's state of health or in premature death. Any treatment received to the right of MS_{a_1} can be attributed to additional medical care over and above that needed to maintain the person's state of health. Consumption of medical services in this area would depend on the consumer's preference map for medical services, and the slope of the price line for medical care.

[5/] Hereafter the words "maintaining the consumer's state of health" will be used synonymously with, and be taken to have the same meaning as, "preserve, improve, or prevent further deterioration of the consumer's (patient's) state of health."

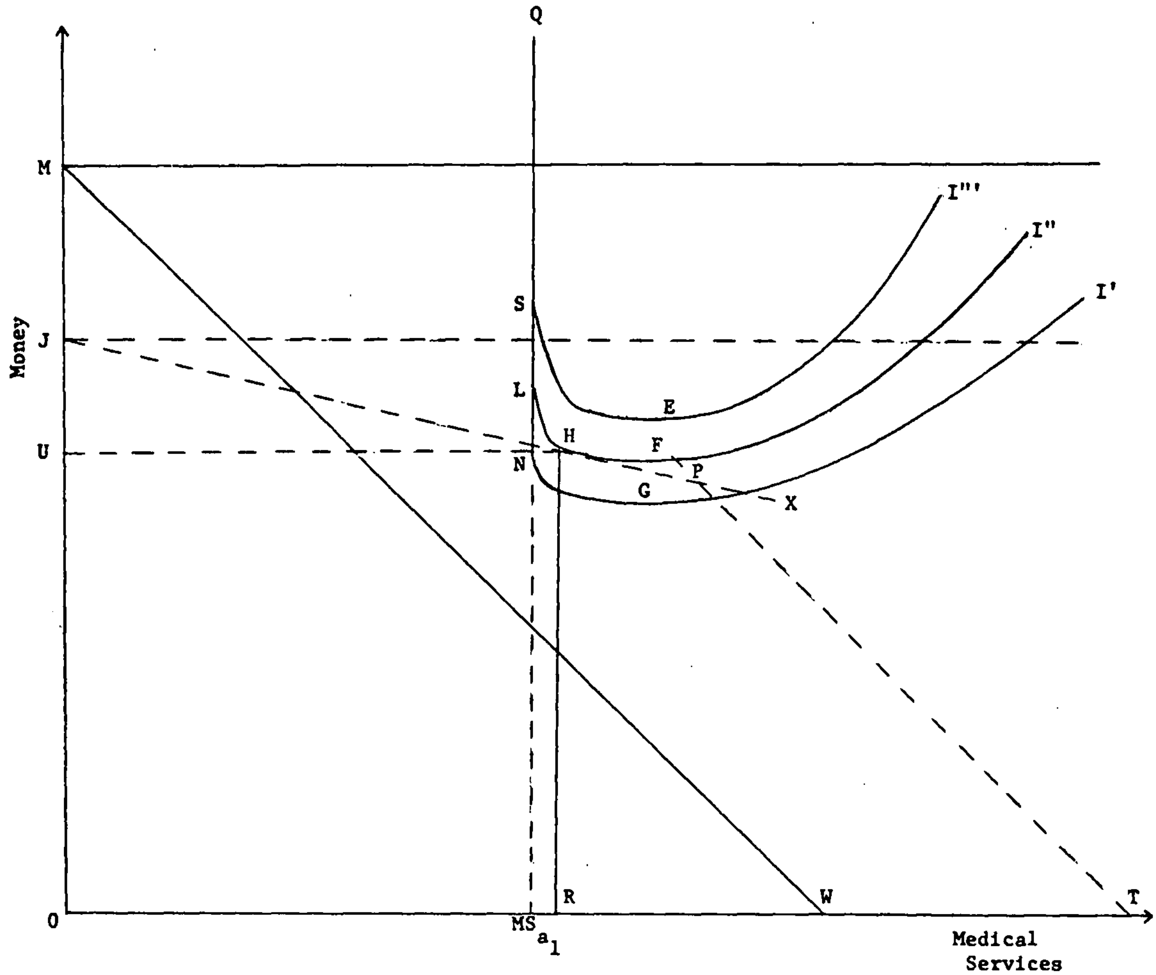

Figure 2. Preference map of an ill consumer given air quality a_1.

The consumer is ill, hence his whole indifference map has changed. QSI''', QLI'', and QNI' are defined as indifference curves in Figure 2. The slopes of these curves are unique. Given the nature of treatment received for the disease, and the alternatives of premature death or continued pain and suffering when treatment is not obtained, a consumer would be willing to give up any amount of money in order to receive treatment. The marginal rate of substitution of medical services for money is infinite in the area O to MS_{a_1}. This implies that holding money alone does not have utility. The amount of medical services necessary to maintain one's state of health must first be obtained. Once MS_{a_1} is reached the consumer's need for medical services has been met. To the right of MS_{a_1} money becomes increasingly more important than the consumption of medical services. The marginal rate of substitution of medical services for money rapidly declines. Hence, as with Figure 1, it does not take too much additional treatment for the consumer to become satiated of medical services. This is represented by E, F, and G on QSI''', QLI'', and QNI', respectively, in Figure 2. To the right of points E, F, and G medical services become a discommodity, and the indifference curves attained a positive slope.[6/]

[6/] It is important to emphasize again that the preference map in Figure 2 is based upon a prescribed set of assumptions. In essence, this map is drawn to explain why a hypothetical consumer would seek medical care. This does not deny the fact a person, Medical services, MS_{a_1}, are prescribed by the physician, which the consumer will at least purchase in his desire to maintain his state of health. The relevant portion of the consumer's budget constraint, with insurance, is JP which is tangent to the indifference curve QLI" at H. The consumer purchases slightly more medical services at R than the recommended MS_{a_1}. At this tangency, utility is maximized and the consumer will pay, in addition to MJ on health insurance, JU for the purchase of R medical services.[7/]

What happens if air quality deteriorates from a_1 to a_2 within time period t? Assuming the same representative consumer to be

---

having a disease affected by air pollution, may not seek medical care and, hence, have a preference map quite different from that described. While such occurrences can result in social costs (through value of production foregone, resulting from absenteeism or reduced efficiency on the job, for example), no demand for medical services is placed upon the system. These incidences would not enter the model. Hence, any estimate of air pollution costs from the model would be an under-estimate of the true costs of air pollution.

There is also the possibility of a consumer seeking medical care who is not ill. Earlier it was necessary to assume the physician would do all that was necessary to preserve or improve his patient's health state. It is now necessary to assume that such abuses as seeking medical care when one is not ill are minimal, and when discovered by the physician would not be tolerated. Given the nature of the Kaiser system, where the emphasis is on providing as efficiently as possible only those medical services which are needed to preserve, improve, or prevent further deterioration of the patient's health state, it is highly unlikely that such occurrences will be a problem.

[7/] Examination of Figure 2 will show the consumer to be better off purchasing health insurance. To see this, construct a line (MW) running through M and parallel to PT. To consume the required medical services MS_{a_1} without insurance would place the consumer on an indifference curve well below QLI".

afflicted with a disease aggravated by air pollution, one would expect a tendency for the patient to become ill more often, and/or the severity of each disease incident to increase. In both instances the result implies an increase in demand for medical services.

Look now at Figure 3. All definitions that applied to Figure 2 are also applicable to Figure 3. In Figure 3, the consumer's entire preference map for medical services has shifted horizontally to the right. Such a shift does represent a change in the consumer's preferences for medical services. Now MS_{a_2} is the quantity of medical services required to maintain the patient's state of health. The consumer's relevant budget constraint is again JP, which is tangent to QNI' at V. The consumer will pay JZ for the consumption of Y medical services, plus MJ for the health insurance premium.

In comparison with Figure 2, MS_{a_1} is less than MS_{a_2} (the amount of medical services necessary to maintain the consumer's state of health has increased); JZ is greater than JU (total monetary outlay for medical services has increased). From a monetary standpoint the consumer is worse off. The increase in the consumption of medical services resulted from the deterioration in air quality, and the resulting increase in monetary outlay for medical services (the difference between JZ and JU) represents an economic cost which could have been foregone if air pollution had not increased.

Two important implications can be drawn from Figures 2 and 3

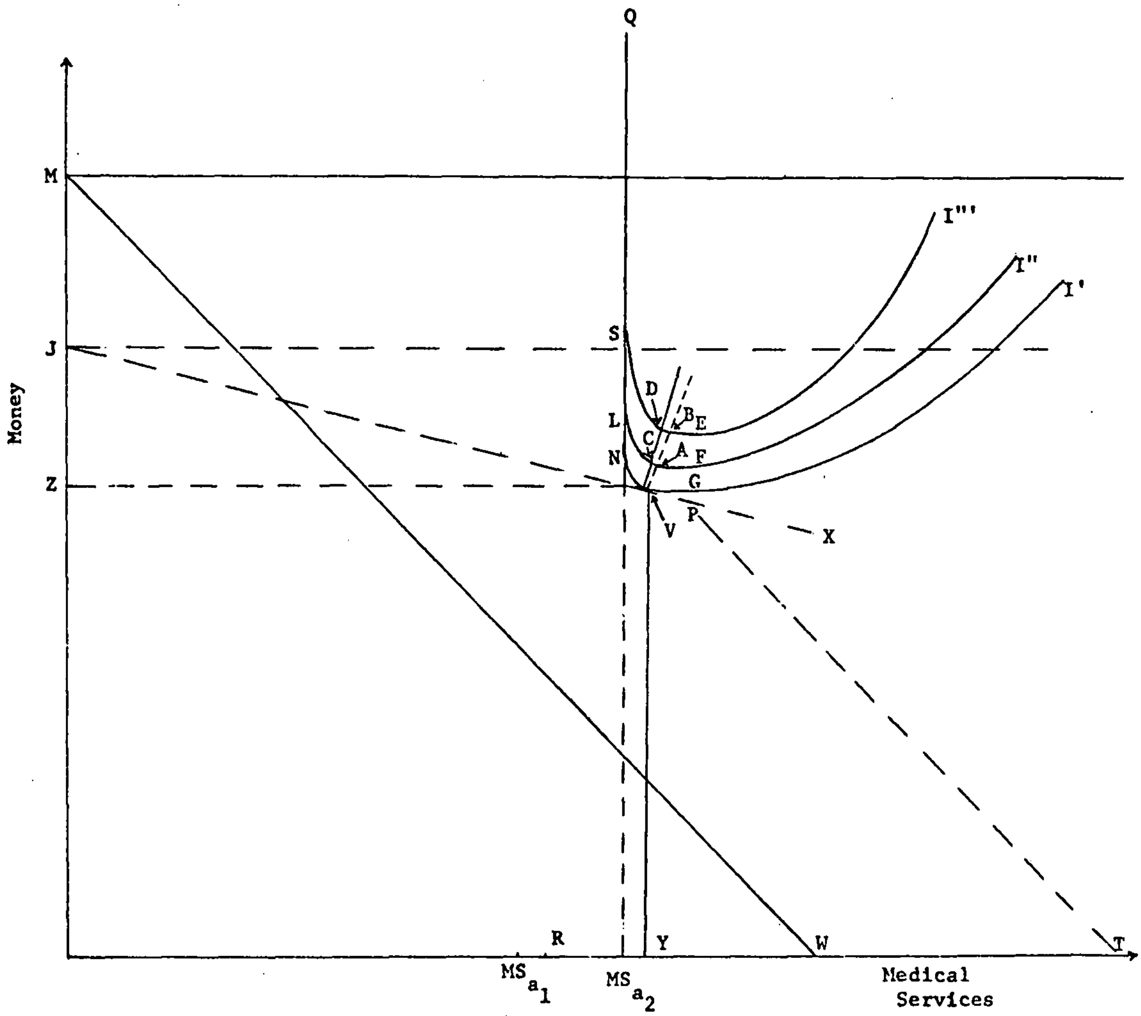

Figure 3. Preference map of an ill consumer given air quality a_2.

with respect to the price and income elacticities of demand for medical services. In Figure 3 the dashed line connecting points VAB represents a price-consumption curve.[8/] The curve is steeply, positively sloped, and in essence shows that as the price of medical services change, the amount of medical services demanded changes proportionately less. The price elasticity of demand for medical services is highly inelastic. Also in Figure 3 the solid line connecting points VCD represents an income consumption curve.[9/] The curve is highly, positively sloped, and illustrates that the amount of medical services demanded will not change materially with changes in the consumer's income. Income elasticity of demand for medical services is close to zero. Each concept (price and income elasticities) willnow be discussed in detail.

The demand for medical services to treat diseases affected

[8/] The price-consumption curve is the locus of utility maximization points achieved when nominal money income and the price of one good are held constant, while the price of the other good is allowed to vary. The locus of equilibrium points is obtained, graphically in Figure 3, by letting JX rotate about J, until a tangency point is obtained with each indifference curve.

[9/] The income consumption curve is the locus of utility maximization points achieved when the nominal money income of the consumer is allowed to vary, and the prices of the two goods, reflected in the indifference map, are held constant. The locus of equilibrium points is obtained, graphically in Figure 3, by drawing other lines parallel to the relevant budget constraint JP and tangent to each indifference curve. The other parallel lines reflect different income levels of the consumer. Connecting the equilibrium points yields the income consumption curve.

by air pollution is assumed to be price inelastic. This assumption follows from the nature of the diseases being examined (without treatment the patient can die a premature death, or experience increased pain and suffering); from the minimum amount of medical treatment required, regardless of price, to maintain the consumer's state of health; from the inability of the consumer to effectively evaluate alternative medical procedures at different prices. All these tend to make the consumer less price-conscious.[10/]

With respect to the Kaiser data, the assumption of price inelastic demand would be particularly relevant. Here the main cost of participating in the system is the insurance premium of a fixed amount. In many instances the premium is paid by the employer. Hence, the cost to the Kaiser patient of using the system, in addition to the insurance premium if he should pay it, is his private costs of time and transportation, and some minimal charges for physician visits and drugs.

[10/] The assumption of price inelastic demand is supported by empirical evidence. Other studies have shown price elasticity of demand for medical services (physician and hospital) to be highly inelastic (18, p. 24-25, 66-67, 72).

Because of the market structure for medical services, most charges for medical services are fairly standardized within a given region. Physicians and hospitals are provided with fee service schedules for various regions in Oregon. While the suggested charges do not have to be followed, they do provide a basis for establishing fees for medical services. (Conversations with Sheldon Wagner, M.D., Head, Oregon State University-Environmental Health Unit, Good Samaritan Hospital, Corvallis, Oregon, July 1971; and see (51, p. 363).

Changes in the consumer's real income will not greatly affect the quantity of medical care purchased (taking into account, of course, other items which can influence expenditures for medical care), i.e., the <u>income elasticity of demand</u> for medical treatment of diseases affected by air pollution is very low. This follows from the assumption of a minimal amount of treatment needed to treat the disease and maintain or improve the patient's state of health.[11/]

The assumption of income inelasticity does not deny the fact that certain amenities of medical care might be a function of income (private room versus a bed in a ward of a hospital, for example). Where possible, such amenities, which, in most instances, really have nothing to do with the quantity or quality of medical care, should be accounted for in determining income elasticity of demand for medical care. At Kaiser, such amenities are non-existent or minimized. The emphasis at Kaiser is to provide the necessary medical services to treat an illness. Conveniences not necessary for such treatment are not provided. For example, amenities such as the use

---

[11/] Empirical evidence on physician services consumed support this assumption to some extent. Andersen and Benham (2) found income elasticities, using permanent income and taking other variables into account (demographic characteristics and preventive care, among others) to be fairly low (less than 0.30). By substituting the quantity of physician services consumed for the monetary outlay (dollar expenditure by the consumers) on the same services, they found income eleasticity to be as low as 0.01 (2, p. 90). The measures of elasticity differ because of free medical care provided at low or no cost to poorer consumers.

of a private room, when a bed in a ward would afford the same medical services, are not obtainable at Bess Kaiser Hospital in Portland.

Medical treatment for diseases associated with air pollution have no close substitutes with other, non-medical, consumption goods; i.e., the cross-price elasticity between medical treatment for these diseases and other goods is zero, or near zero. The acceptance of this assumption rests upon the nature of the diseases being examined. The assumption of zero cross-price elasticity does not negate the fact there might be treatment substitutes within a bundle of treatments used to treat a disease incident. However, once a method of treatment is decided upon a minimum amount is needed, and changes in the prices of other, non-medical goods, will not affect the quantity of treatment purchased. This implies that shifts in the demand for medical care of diseases affected by air pollution will be minimal when the prices of other goods change. Again, this assumption is not inconsistent with the Kaiser data.

With a prepaid medical system, where in many instances the insurance premium is paid by someone other than the patient, the price of medical services and the patient's income play a minimal role in his consumption of services. The assumptions with respect to price, income, and cross elasticities of demand, backed to some extent by empirical evidence from other studies, allow the following argument to be made. The consumption of medical services in the treatment of certain diseases are not affected by the price of the services or the patient's income, regardless of the sources from which care was received. Hence, the use of a prepaid system such as Kaiser's can be used to estimate the medical costs of certain diseases associated with air pollution, and the estimates derived will be representative of other medical systems.

A Broader Framework of Analysis

In order to isolate the effect of air pollution on health, account must be taken of other conditions which could also influence the consumption of medical services. These other conditions can be broadly grouped into socioeconomic-demographic characteristics and meteorological conditions. This section expands the graphical presentation above to specify other variables which might be considered to influence the consumption of medical services and to affect disease incidence rates. This section specifies the general framework within which the variables should be considered, and lays the ground-work for specification of the statistical models used in analysis.[12/]

The Kaiser membership contains N people. On a given day (j)

[12/] What follows has been formulated with the Kaiser data base specifically in mind. With modifications, it can be applied to other data sources.

there is a level of air pollution which is assumed (for illustrative purposes) to be higher than the previous day (j-1). The effects of air pollution on health, and the hypothesized increase in demand for medical services on day j, or shortly thereafter, can show up in two ways. That is, there are two ways the increase in pollution can result in increased demand for medical services. Each provides a supposition about air pollution effects on health. (a) Deterioration in air quality can increase the severity of the disease (i), requiring additional, and perhaps more intense, medical care beyond that which would normally be incurred; (b) increases in air pollution can precipitate an increase in the incidence of certain diseases.

The general framework to be developed for testing the first supposition is specified in Equation (2. 1). With Kaiser, medical services are received at various clincis (outpatient care) or at the hospital (inpatient care and outpatient care). The problem is to determine the portion of total medical costs attributable to air pollution for each disease i for which treatment was sought. The other explanatory variables needed to isolate the effects of air pollution are noted in Equation (2. 1).[13/]

[13/] It is recognized with this model, and all the other models to follow, that there can be lags of several days between the onset of the disease i and exposure to air pollution or meteorological conditions. While the lags are not specified within the definitions of the variables, the potential for them to exist is understood and will be explored in the statistical analysis.

$$I_{ijk} = h(A_{jk}, W_{jk}, S_{ijk}) \qquad (2.1)$$

where

I_{ijk} = an index of medical services consumed (in-patient and out-patient) for treatment of the i^{th} disease (per disease incident or episode) resulting from exposure to air pollution on the j^{th} day for the k^{th} person;

A_{jk} = a measure of air quality on day j for the k^{th} person;

W_{jk} = a measurement of meteorological conditions on day j for the k^{th} person;

S_{ijk} = the socioeconomic-demographic characteristics of the k^{th} person, on the j^{th} day of exposure for the i^{th} disease; and

$(i = 1, \ldots, n)$; $(j = 1, \ldots, m)$; $(k = 1, \ldots, l)$.

Index of Medical Services. An index of medical services (I) would overcome some of the inherent disadvantages incurred by using dollar expenditures for medical services. The index would be based on physical quantities of medical services consumed, would permit aggregation of different kinds of medical services, and would reflect the intensity of medical services used to treat the disease state.[14/] The index, if constructed properly, would more adequately

[14/] The intensity of medical care can be illustrated as follows: For a given disease state there may be several options of treatment open. The option selected is based upon the severity of the disease state. As the disease states become more severe, each option may reflect the quantity of medical services consumed than would dollar expenditures on medical services, because of variations in fees charged for services performed and the availability of free medical care to some patients. An index, with the three characteristics prescribed above, would be more sensitive to changes in air pollution than dollar expenditures.

Air Quality. The possible effects of a deterioration in air quality (A) on the demand for medical services are many. Exposure to most pollutants in sufficient concentrations, and for long enough time periods, can result in harmful physiological effects to the human body. This presentation will give a brief summary of some of these effects.

Particulate matter can effect human health by its chemical composition and size. The particle may be intrinsically toxic due to its innate chemical and/or physical characteristics (75, p. 141). For example, sulfur dioxide when absorbed on particulate matter can impare lung tissue (44, p. 37). Particle sizes less than two or three microns[15/] can penetrate deep into the respiratory system (44, p. 33).

A number of recent laboratory and clinical studies have led

---

represent a more intensive use of medical services to treat the disease as compared to a less severe case of the same disease.

[15/] A micron is approximately 1/25,000 of an inch (77, p. 1).

to a concern that subtle cardiovascular and central nervous system effects may be associated with elevated levels of carbon monoxide in the ambient air (78, p. 9-10). Photochemical air pollution (Los Angeles smog) can cause eye irritation (1, p. 96) and aggravate respiratory diseases (80, p. 9-30). Hydrocarbon air pollutants, which enter into and promote the formation of photochemical smog, have been associated with eye irritation. Aldehydes, also part of photochemical smog reactions, have been shown to irritate the eyes and upper respiratory tract (79, p. 7-27).

Sulfur dioxide, at high enough concentrations, irritates the upper respiratory tract (44, p. 37). A significant correlation has been shown between mortality resulting from chronic bronchitis and sulfur dioxide in England and Wales (44, p. 69).

Meteorological. Meteorological conditions (W) can act independently of, as well as in conuunction with, air pollution to place additional stress on the body, and possibly increase the consumption of medical services. Temperature and humidity have an unquestioned effect on health (8, p. 200). Extremes of temperature can place strain on the body and nervous system; prolonged cold spells can contribute to exhaustion (36); cold temperatures can decrease mucous transport which, by means of the cilia, is one of the main methods of cleansing airborne matter from the lungs (33). Winter, weather fronts, and fog are related to the onset and increased severity of chronic bronchitis (45, p. 11). Exposure to cold, fog, or smog may need to be avoided by asthmatics (45, p. 64). Low temperature, low humidity, a wide range in barometric pressure, and wind have been shown to be associated with the daily prevalence of common cold and cough symptoms (9, p. 524). In another study temperature (on the day the illness started, or a few days prior) seemed to be highly associated with upper respiratory disease incidence (63, p. 738).

Intense radiation can lead to heat stroke; intense cold can increase pulmonary disease (23, p. 80-81). Many physicians believe that shortness of breath in those prone to heart attacks are increased during spells of hot, muggy weather and that fatigue worsens in foggy weather (38, p. XIV). It was found in the Netherlands that the highest mortality from arteriosclerotic heart disease occurs during the coldest months of the year (January and February) (64, p. 507-508).

Meteorological conditions can interact with air pollution to produce an environmental situation which can produce adverse health effects (9). This can happen through chemical reaction of certain pollutants with each other, and with certain environmental phenomena, to produce new, and sometimes unknown substances. These chemical reactions are affected by the presence of wind, sun, and humidity, among other meteorological conditions. The best known example of this is the photochemical reaction between oxides of nitrogen and other organics in the presence of sunlight. The effects on health of the resulting photochemical air pollution, or Los Angeles smog as it is more popularly known, have already been documented above.

Certain meteorological conditions can increase the build up of air pollutants in the environment. Many past air pollution disasters (Meuse Valley, Belgium, 1930; Donora, Pennsylvania, 1948; London, 1952 and 1962, among others) have been characterized by prolonged, anticyclonic high pressure with a secondary inversion (7, p. 2). An inversion in effect acts as a lid, and retains below it all pollution emissions to the atmosphere and effectively retards their dispersion. Hence, air pollutants build up and a minor air pollution problem can become a severe episode if the meteorological conditions under which it was incurred continue.

Chemical reactions of pollutants with each other, the interaction of pollutants with certain meteorological phenomena, and the resulting creation of new pollutants or substantial increase in the amount of pollution, will usually be reflected through adequate air monitoring. However, it is possible that meteorological and air pollution may act interdependently to adversely affect a disease state. That is, the simple air pollution or meteorological condition by itself, and in the absence of the other, may not adversely affect a disease state. However, in combination with each other they may adversely affect the disease condition. Such combined effects would not be reflected by single measurements of different air quality or meteorological phenomena. Hence, in pursuing a study of this type, one should be aware of the possibility of such interdependence.

Socioeconomic-demographic. Other studies have shown that certain socioeconomic-demographic variables (S) play an important role in the incidences of some diseases affected by pollution (37, 39 and 83). It would seem that, to the extent certain variables affect the health of the individual, they should also affect the amount of medical services consumed in treatment of the disease. Included within S are several socioeconomic-demographic variables which would appear to be important in explaining the consumption of medical services and the individual's state of health.

The smoking characteristics of the patient are expected to play an important role in his consumption of medical care. Heavy smoking can aggravate certain respiratory diseases (emphysema and bronchitis, for example), and place additional stress on the body's cardiovascular system.[16/] Hence, the more one smokes, the more medical services one would be expected to consume.

Cessation of cigarette smoking has been shown to be associated

---

[16/] See particularly the following (15, 31, 44, and 45) for the effects of smoking and air pollution on certain respiratory-cardiovascular diseases.

with a decrease in the prevalence of symptoms related to smoking (cough and shortness of breath, for example) (82, p. 146). Numerous other studies have shown the prevalence of cough, sputum production, breathlessness (dyspena), and chronic bronchitis, among other respiratory conditions, to be consistently higher for ex-smokers when compared to nonsmokers, and lower when ex-smokers are compared to cigarette smokers (82, p. 195-205). Other studies reveal that the risk of developing lung cancer decreases with smoking cessation (82, p. 11). Most studies indicate that after ten years of not smoking cigarettes, the ex-smoker's chances of dying a premature death due to a smoking related disease is less than that for a smoker and slightly greater than that for a nonsmoker.[17/] These studies indicate that account should also be taken of past cigarette smokers. The effects of past smoking can still aggravate certain respiratory diseases, and increase the consumption of medical services. The more recent the ex-smoker quit smoking, the greater the potential for consuming more medical services. The studies would also seem to indicate that there are certain long-term effects of cigarette smoking, some of which are irreversible (82, p. 145), which would make the ex-smoker intrinsically different from the nonsmoker.

[17/] Obtained from the American Cancer Society, and based on their slide series "The Time to Stop is Now."

The occupation of the individual could affect the consumption of medical care. It is known that exposure to dusts can aggrevate asthma (31, p. 34) and chronic bronchitis (45, p. 13). Hence, it would seem that dustier occupations could increase the consumption of medical services for individuals with a respiratory ailment.

The residential history of the patient is also important. Studies have shown that the incidence and death rates from emphysema and cancer are higher in the cities than in the relatively non-polluted rural areas (44, p. 71-72). It is argued that an individual's state of health would be better if he had lived most of his life in a rural area or other unpolluted environment. The writer feels that, if this were the case, when the patient does become ill his disease state would be less chronic, and require less consumption of medical services.

The incidence rates of certain diseases are more prevalent in certain age, race, and sex groups (15, p. 13; 45, p. 39). Hence, it is felt race, age, and sex are three variables which could influence the demand for medical services by an ill person. If some diseases are more predominant in certain age, race, and sex groups, then persons within these groups having the diseases might be more acutely ill than individuals outside the groups who might also contract the disease. Also, with increasing age, the efficiency of the lung function declines, due probably to a decrease in elasticity of lung tissue, increasing inflexibility of the chest wall, and perhaps to impairment of the regulation of ventilation and of the efficacy of cough (15, p. 16). These aging lung conditions would tend to impair the clearance of pollutants from the respiratory tract, which would result in increased aggravation of respiratory diseases and an increase in the consumption of medical services.[18/] Also, increased aging could be one of the main causes of chronic obstructive pulmonary diseases (15, p. 16).[19/] One would also expect an older person to generally have a poorer state of health, which could result in increased consumption of medical service per disease incident, as compared to a younger person contracting the same disease.[20/]

Another social habit which might affect a patient's consumption of medical services is his drinking habits. Excessive drinking can affect a person's state of health (emotionally, mentally, and physically). Alcoholism can play a role in pumonary tuberculosis (45, p. 32). Excessive consumption of alcohol leaves a person's body in such a physical state that, if he did become ill, the consumption

---

[18/] Conversation with Sheldon Wagner, M. D., August 25, 1971.

[19/] Chronic obstructive pulmonary disease is a term which applies to those patients with chronic bronchitis, asthma, and anatomic emphysema, who exhibit persistent obstruction of bronchial flow (15, p. 7).

[20/] Sheldon Wagner, M. D., August 25, 1971.

of medical services could be increased. Alcoholism has been extensively associated with respiratory problems.[21/]

It is felt that the more physically active a person is, the better his state of health would be. This could imply that when a person does become ill, the disease case could be milder, which would mean less medical services required in the treatment of the disease incident.[22/] There is a suspected association between lack of physical excercise and coronary heart disease (25, p. 15).

The number of persons per room within the patient's home would be an indicator of the crowding within the patient's environment. Crowded conditions could result in increased consumption of medical care. More crowded conditions lead to increased physical and emotional strain being placed upon the individual, the increased possibility of initial infection with certain respiratory diseases (tuberculosis, for example) (45, p. 31), and the increased chance of possible reinfection.

The permanent income of the patient is included in an attempt to describe the consumer's ability to make expenditures, in lieu of medical expenditures or services, preserving and bettering his state of health. Examples of such expenditures would be those made for

---

[21/] Sheldon Wagner, M.D., August 25, 1971.

[22/] Sheldon Wagner, M. D., August 25, 1971.

sound housing and nutritional food. To the extent the patient's health may deteriorate because he is not able to make such expenditures on himself or his family, the more medical services one would expect to be consumed. Higher values of permanent income are expected, with the Kaiser data base, to decrease the consumption of medical services.

The marital status of the patient could explain some consumption of medical services. Available evidence indicates the tendency of single persons to consume more total days in the hospital than married persons (18, p. 60). Several reasons could be advanced for this. If a patient is married, assuming the marriage to be a stable one, he would usually have better nutrition, a more regular schedule, and, overall, would be better physically. This could tend to contribute to a better state of health for the married versus the single person.

It has been suggested that another possible reason for married persons to consume less total hospital days than single patients are the availability of persons in the patient's home to take care of him (18, p. 60). This may be true, but single patients can also have persons available to take care of them when they are ill. A better approach to this concept would be to measure it directly. Hence, another explanatory variable for the consumption of medical services would be a measure of the availability of someone at home to take care of the ill patient. However, the variable would probably be more important in explaining the consumption of inpatient medical services than outpatient medical services. Once a person goes to the clinic, he will most likely be treated in the prescribed way.

The general framework specified in Equation (2. 1) would afford a test whether air pollution increases can cause greater utilization of medical services. To determine whether deterioration in air quality can precipitate an increase in the incidence of certain diseases requires another general model illustrated by Equation (2. 2).

$$\frac{n_{icj}}{N_{cj}} = t(\bar{A}_{cj}, \bar{W}_{cj}, \bar{S}_{cj}) \tag{2. 2}$$

where

n_{icj} = the number of people in the geographical area c making new or initial contact with the medical system for disease i on day j;

N_{cj} = the total number of people living in geographical area c on day j;

$\bar{A}_{cj}$ = an average measure of air quality over the geographical area c on day j;

$\bar{W}_{cj}$ = an average measure of meteorological conditions over the geographical area c on day j;

$\bar{S}_{cj}$ = an average of socioeconomic-demographic characteristics of people living within geographical area c on the j^{th} day; and

($c = 1, \ldots, p$); ($i = 1, \ldots, n$); ($j = 1, \ldots, m$).

The independent variables in Equation (2. 2) are expressed as a mean over a geographical area (census tract, for example). The approach outlined in (2. 2) offers one method of seeing whether differences in air quality, meteorological, and socioeconomic-demographic conditions can play a role in people contracting disease i.

The arguments for including the explanatory variables $\bar{A}_{cj}$, $\bar{W}_{cj}$, and $\bar{S}_{cj}$ in Equation (2. 2) are similar to the arguments given for their inclusion in Equation (2. 1). The principal influence of $\bar{A}_{cj}$, $\bar{W}_{cj}$, and $\bar{S}_{cj}$ in Equation (2. 2) is that they affect a state of health, hence making the patient more susceptible to certain cardiovascular-respiratory problems. Therefore, $\bar{A}_{cj}$, $\bar{W}_{cj}$, and $\bar{S}_{cj}$ are expected to have the same causal relationships on disease incidences as A_{jk}, W_{jk}, and S_{ijk} had in the consumption of medical services in Equation (2. 1). The exceptions to this are discussed below.

Several additional variables are included in $\bar{S}_{cj}$ which were not considered in S_{ijk}. These variables would be expected to influence disease incidence ($\frac{n_{icj}}{N_{cj}}$), but would not necessarily influence the consumption of medical services (I_{ijk}). The variables are reading of medical literature, distance to the nearest clinic or hsopital, and a measure of the hypochondriac tendencies of the patient to contact the medical system.

The reading of medical literature could explain differential information and preferences toward recognizing the need for, and in appreciating the desirability of, seeking medical care. This would, in many instances, imply contact with the medical system to provide medical services. Even if medical services were not provided the patient, contact with the system, where a disease is positively identified, would constitute a disease incident. It is expected that increased education in the medical area will be positively related to disease incident, particularly with respect to patients having the cardiovascular-respiratory problems with which this study is concerned.

Distance to the nearest hospital or clinic reflects, in part, the out-of-pocket costs and inconvenience to the potential patient of obtaining medical care from Kaiser. Since all Kaiser clinics are located within the immediate Portland area[23/] and the hospital is located in North Portland, it is quite possible that potential patients living in outlying areas would substitute medical services closer to

[23/] One clinic is in Vancouver and one is in Beaverton (see Figure 4 in Chapter III), the other three clinics are located within the Portland City Limits.

place of residence for Kaiser services. Longer distances could delay the patient seeing the Kaiser doctor, or not seeing him as often. Each situation could have direct bearing on the incidence rate. Hence, it is expected that the further patients have to go in obtaining medical services from Kaiser, the lower will be the disease incidence as reflected in the Kaiser data.[24/]

The medical hypochondriac has a morbid fear or anxiety about his health state. As explained in the last section, it is highly unlikely that such occurrences will be a problem as they affect the actual consumption of medical services. However, to the extent the potential patient would telephone or visit the clinic or hospital, he would positively affect the disease incident rate. That is, the more prone patients are to do this, the higher the incidence rates will be. Hence, a measure of the tendency of a patient to contact the system at the first sign of a perceived illness should be included within the analysis.

The framework specified in Equation (2.2) would afford a test whether air pollution can precipitate an increase in the incidence of certain diseases. The next section will specify the null hypotheses to be tested, and delineate the statistical models by which they will be tested.

[24/] This would imply that as the out-of-pocket costs of using Kaiser facilities increase with distance, the demand for Kaiser medical services becomes more elastic. However, demand for medical services as a whole would still remain inelastic because medical services from other facilities closer to the patient could be substituted for the Kaiser services.

Specification of the Statistical Models

This section specifies the statistical models, and the hypothesis to be tested by each. The socioeconomic-demographic and medical data used in the equations come from a random sample of the total Kaiser membership. Only outpatient medical records, and the socioeconomic data of the sample members responding to a questionnaire, administered to them during 1970 and 1971, were included in this study. The collection of these and other data, the selection of what data to use and their preparation for use in the models, and the process of deciding what variables to include within the models will be detailed in the next several chapters.

The first null hypothesis to be tested is as follows:

Ho_1 : Deterioration in air quality causes no increase in the consumption of medical services per outpatient contact with the medical system.

Equation (2. 3) permits the testing of this hypothesis.

$$Y_{ijk} = \beta_0 + \beta_1 X_{1ijk} + \beta_2 X_{2ijk} + \beta_3 X_{3ijk} + \beta_4 X_{4ijk} + \beta_5 X_{5ijk} + \beta_6 X_{6ijk} + \beta_7 X_{7ijk} + \beta_8 X_{8ijk} + \beta_9 X_{9ijk} + \beta_{10} X_{10ijk} + \beta_{11} X_{11ijk} + \beta_{12} X_{12ijk} + \beta_{13} X_{13ijk} + \beta_{14} X_{14ijk} + \epsilon \quad (2.3)$$

where

Y_{ijk} = index of consumed outpatient medical service for disease i resulting from contact with the Kaiser system on day j by person k, expressed in dollars.

X_{1ijk} = age of the k^{th} person on day j who consumed outpatient medical services for disease i.

X_{2ijk} = sex of the k^{th} person who consumed outpatient medical services on day j for disease i.

X_{3ijk} = marital status of the k^{th} person who consumed outpatient medical services for disease i on day j.

X_{4ijk} = number of people in the k^{th} patient's household who consumed outpatient medical services for disease i on day j.

X_{5ijk} = household income of the k^{th} patient who consumed outpatient medical services for disease i on day j.

X_{6ijk} = race (Negro) of the k^{th} person who consumed outpatient medical services for disease i on day j (one if Negro, zero otherwise).

X_{7ijk} = race (Others) of the k^{th} person who consumed outpatient medical services for disease i on day j (one if of another non-white race, zero otherwise).

X_{8ijk} = a great deal of time and energy expended by the k^{th} patient, who consumed outpatient medical services for disease i on day j, in being physically fit (one if a great deal of time and energy expended, zero otherwise).

X_{9ijk} = some time and energy expended by the k^{th} patient, who consumed outpatient medical services for disease i on day j, in being physically fit (one if some time and energy expended, zero otherwise).

X_{10ijk} = a measure on the frequency of alcohol consumption by the k^{th} patient who is a heavy drinker, and consumes outpatient medical services for disease i on day j.

X_{11ijk} = an index expressing the cigarette smoking characteristics of the k^{th} patient consuming outpatient medical services for disease i on day j.

X_{12ijk} = an index of occupational exposure to job related pollutants by the k^{th} patient consuming outpatient medical services for disease i on day j.

X_{13ijk} = a measure of the ambient air pollution that the k^{th} patient, who consumes outpatient medical services for disease i is exposed to on day j (expressed as suspended particulates, micrograms per cubic meter per day).

X_{14ijk} = a measure of the meteorological conditions that the k^{th} patient, who consumes outpatient medical services for disease i, is exposed to on day j (expressed as degree days - the absolute difference between the average daily temperature and $65^{o}F$).

ϵ = a random error term, where $\epsilon \sim N(0, \bar{\sigma}^2)$,[25/] and (i = 1, . . . , n); (j = 1, . . . , m), (k = 1, . . . , l). And $\beta_0, \ldots, \beta_{14}$ are the parameters to be estimated.

The second hypothesis to be tested by this study was as follows:

Ho_2 : Deterioration in air quality does not precipitate an increase in the number of contacts with the medical system per disease category.

While it was desired initially to test whether air pollution precipitated an increase in the incidence of disease i, incidence was discarded as a potential dependent variable after further investigation. The reason for this was as follows. Incidence is defined as an expression of the rate with which a certain event occurs, such as the number of <u>new</u> cases of a specific disease occurring during a certain time period (16, p. 730). The problem with incidence is

---

[25/] See (17, p. 17) for a detailed presentation about the usual assumptions concerning the random error term.

that the disease must be a new case before it is measured, i.e., the disease must be a new episode. Many chronic and acute respiratory diseases can be part of a continuing episode, and be affected by air pollution such as to precipitate additional visits and/or contacts with the medical system. Use of the incidence would not reflect these additional contacts, and hence would inadequately express demands placed on the medical system as they might be affected by air pollution.

The second hypothesis cannot be directly tested, however, because of an error in the specification of the statistical model. The dependent and independent variables were specified only for sample members who contacted the Kaiser system per census tract. The total number of people in the sample at risk per census tract were not included. Hence, a test of the model, specified in this format, was meaningless. The computer and programming expense to correct the error was prohibitive.

However, there did appear to be an alternative method of specifying the model which provided the possibility of testing the hypothesis indirectly. While the socioeconomic-demographic characteristics of the population at risk could not be determined easily, the number of people at risk per census tract was known. Hence, the alternative model (2.5) was specified. Within its dependent variable are included the number of contacts per disease category which resulted in the consumption of medical services per census tract, and the number of people at risk per census tract.[26] The indirect test of the second hypothesis will become clearer with specification of the alternative model.

The dependent variable (Y_{icj}) of Model (2.5) is defined by (2.4) as follows:

$$Y_{icj} = \frac{\sum_{k=1}^{1} Y_{icjk}}{N_{cj}} . \qquad (2.4)$$

where

$\sum_{k=1}^{1} Y_{icjk}$ = the sum of the index of consumed outpatient medical services for disease i for all k persons in census tract c on day j.

N_{cj} = the total number of Kaiser respondents residing in census tract c on day j; and

$(i = 1, \ldots, n)$; $(c = 1, \ldots, p)$; $(j = 1, \ldots, m)$; $(k = 1, \ldots, 1)$

This leads to the specification of the statistical model (2.5).

---

[26] As will be discussed in Chapter VIII, even if the model to test the second hypothesis had been specified correctly, it could not have been tested because there were not enough observation per census tract.

$$Y_{icj} = \beta_0 + \beta_1\overline{X}_{1icj} + \beta_2\overline{X}_{2icj} + \beta_3\overline{X}_{3icj} + \beta_4\overline{X}_{4icj} + \beta_5\overline{X}_{5icj} + \beta_6\overline{X}_{6icj} + \beta_7\overline{X}_{7icj} + \beta_8\overline{X}_{8icj} + \beta_9\overline{X}_{9icj} + \beta_{10}\overline{X}_{10icj} + \beta_{11}\overline{X}_{11icj} + \epsilon.,\qquad (2.5)$$

where

$\overline{X}_{1icj}$ = the average age of patients consuming outpatient medical services for disease i from census tract c on day j.

$\overline{X}_{2icj}$ = the percentage of the female sex consuming outpatient medical services for disease i from census tract c on day j.

$\overline{X}_{3icj}$ = the percentage of married patients consuming outpatient medical services for disease i from census tract c on day j.

$\overline{X}_{4icj}$ = the average number of people per household in census tract c for individuals consuming outpatient medical services for disease i on day j.

$\overline{X}_{5icj}$ = the average household income in census tract c for individuals consuming outpatient medical services for disease i on day j.

$\overline{X}_{6icj}$ = the percentage of none-white patients in census trace c consuming outpatient medical services for disease i on day j.

$\overline{X}_{7icj}$ = the average frequency of alcohol consumption by heavy drinkers in census tract c who consume outpatient medical services for disease i on day j.

$\overline{X}_{8icj}$ = the average smoking index of patients in census tract c who consume outpatient medical services for disease i on day j.

$\overline{X}_{9icj}$ = an average index of occupational exposure to job related pollutants for patients from census tract c consuming outpatient medical services for disease i on day j.

$\overline{X}_{10icj}$ = the average measure of meteorological conditions for patients consuming outpatient medical services for disease i from census tract c on day j (expressed as degree days - the absolute differences between the average daily temperature and 65^{o}F).

$\overline{X}_{11icj}$ = an average measure of ambient air pollution for patients consuming outpatient medical services for disease i from census tract c on day j (expressed in suspended particulate, micrograms per cubit meter per day).

ϵ = a random error term, where $\epsilon \sim N(0, \sigma^2)$, and

(i = 1, . . . , n); (j = 1, . . . , m) (c = 1, . . . , p). And β_0 and β_{11} are the parameters to be estimated. Y_{icj} is as previously defined in (2.4) above.

The dependent variable in (2.5) is expressed as the per capita cost of outpatient visits over census tract c on day j. Three separate items can influence the dependent variable in (2.5): The consumption of outpatient medical services per visit (Y_{ijk} in (2.3)); the number of contacts with the Kaiser system which result in medical services being consumed; and the number of people at risk per census tract. An increase, resulting from air pollution, in the number of contacts consuming medical services could cause the per capita costs of outpatient medical visits (Y_{icj}) to increase, the number of people at risk assumed constant. Similarly, increased consumption of medical services, resulting or associated with air pollution, would also increase (Y_{icj}). It is highly probable that both of these affects operate together.

Assuming a significant association between air pollution and the per capita costs of outpatient visits in (2.5), one could not determine whether the association would be caused by an increase in the number of contacts with the medical system or an increase in the consumption of medical services per contact. However, the null hypothesis associated with (2.3) permits a test on the consumption of outpatient medical services per visit and air pollution. If this

III. A DESCRIPTION OF THE STUDY AREA AND THE KAISER PLAN

Introduction

This chapter provides a basis for discussing in the next several chapters the data for each component of the study's data system - medical, socioeconomic-demographic, air pollution, and meteorological - and their inclusion within the statistical models. The first section of the chapter describes the study area, and delineates some of the characteristics which are common to the area. The second section presents a description of the overall Kaiser system, and then proceeds to discuss the Oregon Region Kaiser Foundation Health Plan. The Kaiser five percent, ongoing, random sample, from which the medical data used in this study were obtained, is then detailed. The household interview administered to each eligible subscriber and spouse of the five percent sample during 1970 and 1971 is presented. The last section discusses some aspects of the data which are held to be important.

The Study Area

The area of study centers within the Portland Standard Metropolitan Statistical Area (S. M. S. A.) population 1, 009, 129 (68, p. 39-175). The S. M. S. A. is made-up of Clackamas (county seat Oregon

City), Multnomah (county seat Portland), and Washington (county seat Hillsboro) counties in Oregon, and Clark county (county seat Vancouver) in the State of Washington. The study area is shown in Figure 4, and is collectively made-up of the cities designated on the map and the immediate, surrounding non-urban areas. The study area boundaries were determined by the availability of air pollution data and proximity to the Kaiser medical facilities in Portland, Beaverton, and Vancouver. The study period is 1969 to 1970.

Portland, Oregon, population 381,927 (68, p. 39-175), is the largest city in the S. M. S. A. and in Oregon. The city is located in the northwestern part of the state, astride the Willamette River near its confluence with the Columbia River. Together with the rest of the S. M. S. A., the city is an important trading, transportation, and manufacturing center for the Pacific Northwest. Racially the area is mostly white, with small groups of minority races, mostly Orientals and Negroes (57, p. 394).

Industrially Portland is economically diversified. There are a number of branch plants of Eastern companies (57, p. 395). However, Portland does not have the industrial concentration of the Atlantic Seaboard. Portland has seven major industries. These include food processing and related agricultural production; textiles; lumber and wood products, including pulp and paper; chemical;

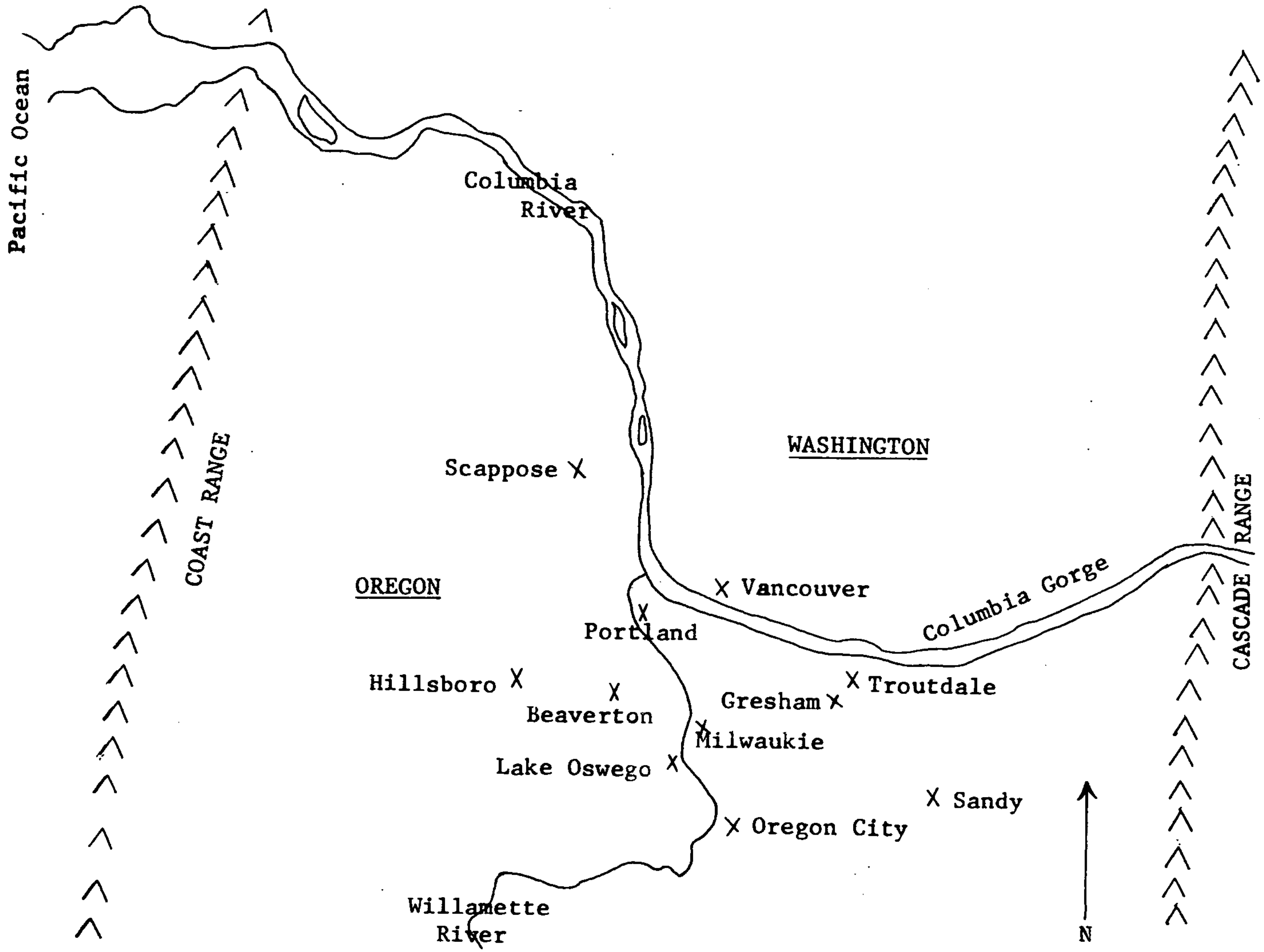

Figure 4. The study area.

aluminum and other metal fabricating plants; shipping (57, p. 395) and tourism. Situated on the Columbia and Willamette Rivers, Portland is an important export center for the Pacific Northwest. This is particularly true in the shipping of grain products from the Inland Empire of eastern Oregon and Washington, all of Idaho, western Montana and Wyoming, and parts of Utah and Nevada (57, p. 395). Tourist visitation is particularly heavy in Portland during the summer, because of Portland's convenient location with respect to many recreational opportunities (73) and its annual rose parade held in early June.

Portland is enclosed by two mountain ranges: the low Coast Range on the west, and the Cascade Range to the east. Each range is approximately 30 miles from the city. Portland is about 60 miles from the Pacific Ocean. The Cascade Range provides a steep high slope for the orographic lift of moisture laden westerly winds and consequent heavy rainfall on the west side of the range. The Cascade Range provides a barrier containing the Interior Columbia Basin with its continental air masses. While most air flow in the Portland area usually is northwesterly in spring and summer, and southeasterly in fall and winter, there are occasional movements of dry continental air from the east to the west through the Cascade passes and the Columbia Gorge. When this occurs, extreme high and low temperatures are experienced in the Portland area (73).

With the exception of these temperature extremes, Portland has a definite mild winter rainfall climate. Average annual temperature is 52.9°F. Annual rainfall averages 37.18 inches with 88 percent of the annual total occuring between the months of October and May. On the average there are only five days each year with measurable snow. Incursions of marine tempered air are a moderating influence on any temperature extremes, with extremes of heat or cold lasting only a few days before being broken (73).

The region's greatest, area wide, pollution problem is suspended particulates (4, p. 4). Periodic inversions during the fall and winter months (September to February) can trap these particulates for several days at a time causing visibility and other problems. Sulfur dioxide, photochemical smog, and nitrogen dioxide are not presently a serious problem in the Portland area (4, p. 3-4). In many instances cloud cover rules out a photochemical smog problem. Primarily only downtown Portland has a problem with pollutant gases, mainly carbon monoxide. While there is some heavy industry in Portland, pollution problems from these sources are usually localized. The suspended particulate problem also extends southward from Portland to the entire Willamette Valley. Extended periods of atmospheric stability have allowed the build-up of pollutants within the Valley (48, p. 12). The wind flow patterns in the Valley are somewhat simpler than they are within the immediate Portland-

Vancouver area, primarily because there is no influence in the Valley of air flow pattern resulting from the Columbia River Gorge. Air flow patterns within the Willamette Valley are primarily channeled by the Coast and Cascade Range: winds are predominately from the north in the summer and from the south in the winter. The sea breezes and associated marine air incursion are of importance in the Valley as they are in the Portland-Vancouver area (47, p. 159).

An Overall View of the Kaiser Medical System

The Kaiser System

The Kaiser-Permanente Medical Care Program is a non-profit, group insurance plan which provides comprehensive medical care to its members. Kaiser is one of the oldest and largest of its kind in the nation. It has 2.2 million members, and was started by Henry J. Kaiser, founder of the Kaiser Industrial empire, in 1938 to provide medical care to construction workers at the remote Grand Coulee Dam site in Washington (62, p. 1).

The plan operates in California, Oregon, Hawaii, Colorado, and Ohio. It has 2,100 doctors, which represent nearly every medical speciality (62, p. 1). The group practice concept of providing medical care is at the heart of the Kaiser Plan. All of the doctors within the six geographic regions independently contract with Kaiser

Foundation Health Plan to provide medical services. Each region develops its own budget. Membership fees are established by region; however, variations among the regions are not considered substantial (62, p. 21).

The physicians are organized into a medical group, a partnership, from which they receive their salaries. A new physician, after three years of straight salary, will be invited to join the partnership. This will usually entitle him to a year end bonus, the amount of which is determined by the success of the region in operating within its budget. The monetary compensation is usually competitative with what the physician would earn in private practice (62, p. 21).

Kaiser's emphasis in providing medical services has been to eliminate unnecessary health care[27] and to concentrate on outpatient care. While the national average for hospital admissions averaged 137.9 a year per 1,000 people, Kaiser overall had cut this to 80 admissions per 1,000 per year. The average length of hospitalization was 6.65 days for Kaiser members compared to 7.8 days for

---

[27] Unnecessary health care, from Kaiser's viewpoint, is providing medical services to the patient which are of little medical value to him, or the provision of the services within a framework which is inefficient. For example, the use of a private room in a hospital when a room of multiple occupancy would serve just as well; or the administering of the services on an inpatient basis when an outpatient visit to a clinic would accomplish just as much, and be less expensive.

the national average.[28] Because he has a financial stake in the efficiency and profitability of the system, the Kaiser physician has an incentive to avoid unnecessary hospitalization (62, p. 21).

In 1970 Kaiser's 2.2 million members made 7.6 million visits to the doctors' offices and spent almost one million days in the hospital. The members were cared for in the Plans' 21 hospitals and 54 medical clinics.

The Oregon Region, Kaiser Foundation Health Plan was established in 1943. The Plan provides prepaid, comprehensive medical care to approximately 146,000 members in 1970.[29] The plan is sold mainly to organized groups (unions). The services are provided on an inpatient and outpatient basis. They include physicians' services in the hospital, or office; hospitalization, surgery, laboratory and X-ray, special nursing, emergency service, mental health, and routine physical examinations, among other medical services.

The membership of the Oregon Region, from a socioeconomic

---

[28] The Kaiser Foundation Health Plan, Oregon Region average length of stay for hospitalization of health plan members was 5.2 days in 1970; average hospital admission per 1,000 members were 87. (Information obtained from Kaiser Foundation Health Plan, Oregon Region, provided by Marilyn McCabe, Special Research Assistant; Kaiser Health Services Research Center; Portland, Oregon.)

[29] This was the membership for the Oregon Region as of December 31, 1970. The information was obtained from Marilyn McCabe, Kaiser Health Services Research Center.

standpoint, is quite diverse. There is a complete span of occupational-socioeconomic groups represented within the membership (20, p. 938). It has been estimated that 15 percent of Portland-Vancouver residents are enrolled in the plan (22, p. 298-299), and receive most of their medical care through Kaiser's hospital and six clinics. Kaiser also provides medical services to non-members on a fee for service basis.

The Permanente Clinic is the partnership of physicians contracting with the Kaiser Foundation Health Plan to provide medical services to the membership. The Permanente Clinic is paid a negotiated fee per health plan member. The fee is paid whether the member seeks medical care or not. The income of the Permanente Clinic is then distributed to the member physicians in a manner determined by them (20, p. 938).

The socioeconomic-demographic and medical data for this research came from the Kaiser Health Services Research Center. Kaiser Research is a separate arm within the Kaiser system. The Center is federally funded to provide a base of funds to attract qualified people of different professional backgrounds to do research on health services delivery as opposed to clinical research. Some of the subjects Kaiser was investigating at the time of this study included the impact of providing new kinds of medical services to a population; impact on utilization, cost, and feasibility of new ways of providing existing services; providing medical care services to new kinds of population; the system properties of the medical care system and making efficient utilization of that system.

There were several advantages in using the Kaiser system for this research. (1) There were approximately 150,000 members in the system. This provided many data for research purposes. (2) The population base at risk for medical services was continually known. (3) The Kaiser system provided a total range of medical services from mental health to drugs. (4) The Kaiser data system was readily amenable to use by modern data processing equipment.

A Description of the Five Percent Sample

The objectives and intent of studies conducted by Kaiser Research were to explain the determinants of medical care utilization. To accomplish these goals a five percent, ongoing, random sample was drawn from the total membership in the Fall of 1966. A comprehensive system for computerizing the medical information for the sample was developed. It was from the medical research on this sample that the outpatient medical data for this study were obtained. This section discusses the drawing of the sample and its socioeconomic-demographic characteristics.

All the membership records of Kaiser Health Plan were recorded on magnetic tape for ease of computer processing. Each family, and member within that family, were identified by a unique number called the Health Plan Identification number (H. P. I. D.).[30/] This provided a reliable sampling framework from which to draw a random, five percent sample of family units. The sampling method approximated a two-stage probability sample. The primary sampling unit was a five percent random sample of the family units. Approximating primary families,[31/] the family units provided natural clusters of individual elements. Kaiser felt that many medical care phenomena were family oriented, hence all the elements (individuals) within the cluster were included. Kaiser described this procedure as being equivalent to subsampling with a sampling fraction of one in the second stage.[32/] The method provided an equal probability cluster sample of the Kaiser Health Plan total population, allowing estimates of medical care utilization to be made on either a family or an individual basis when the appropriate variance formula is used (22, p. 299).

[30/] An H. P. I. D. number was assigned to each Kaiser member. It was a ten place number: the first eight digits identified the family unit, and the last two digits identified members within the family.

[31/] A subscriber family may not necessarily include all the members of the family; that is, some members of the primary family may not be members of Kaiser Health Plan.

[32/] See the following citation for a more complete presentation of this sampling procedure, (27).

The original sample was drawn from a list of people eligible for Kaiser Health Plan service on September 1, 1966. Each month since then a five percent, random sample of all new families to the Health Plan have been added to the overall sample (22, p. 300). Hence, in the aggregate the sample is actually larger than five percent of the total membership.

Once an individual is chosen for the five percent sample, he will always be a member of the sample even if he should become a non-member of the Health Plan. Continuous medical care utilization is recorded whenever members of the sample use or contact the Kaiser System. The original sample had 1487 member family units with 4,123 total individuals. By July 1968, 2,311 families and 6,514 individuals were in the sample (22, p. 300).

The Kaiser five percent sample was designed to be representative of the Kaiser membership at any point in time (22, p. 300). Characteristics of the total membership can be inferred from the sample population. During 1967, 96 percent of the sample members resided within the Portland Standard Metropolitan Statistical Area (S. M. S. A.). Sixty-four percent of the total members resided in Multnomah County (22, p. 302). There has been no reason to believe that the distribution has changed materially for the other years of the study period.

The representativeness of the total Kaiser membership to the

S. M. S. A. is an unanswered question. One Kaiser researcher has gone so far to state that, while not conducting a thorough investigation, the preliminary data he had seen did not indicate any significant differences (61, p. 125, footnote no. 15). Some comparisons of the Kaiser membership to the population of the S. M. S. A. have been made based on the 1960 census. Four percent of the total membership were nonwhite compared with three percent of the S. M. S. A. The Kaiser Health Plan membership had slightly more members in the under 25 age group and had slightly less in the over 65 age group than the S. M. S. A. (22, p. 302).[33/] Dr. Merwyn Greenlick, Director of Kaiser Health Services Research Center, felt that the Kaiser membership was not significantly different from the Portland S. M. S. A. However, since the Kaiser membership was not randomly drawn from the S. M. S. A., he was wary of making inferences from membership or sample results to the S. M. S. A.[34/]

[33/] Some other general characteristics of the Kaiser five percent sample membership, not in comparison with the S. M. S. A., showed that 81 percent were Protestants, and 16 percent were Catholic. Eighty-one percent of the sample subscribers were married. Eighty-seven percent of the subscribers were employed, with semiskilled (24 percent) and clerical sales (23 percent) appearing to be the main occupational categories (22, p. 302).

[34/] Conversation with Merwyn R. Greenlick, Ph. D.; Director, Kaiser Health Services Research Center; Portland, Oregon; November 22, 1971.

The Kaiser Household Interview

One of the specific aims of Kaiser Research was to examine associations between socioeconomic-situational background characteristics and the consumption of medical services in different disease states. It was hypothesized that different background characteristics were important determinants of medical consumption with different disease situations (22, p. 299). To accomplish this objective a household interview survey, to elicit socioeconomic-demographic and other data, was administered to members of the five percent sample. It was from this survey that the socioeconomic-demographic data were obtained for use in this study. This section will briefly describe the purpose of the interview and how it was administered.

A questionnaire survey was administered to each eligible Kaiser Health Plan subscriber, or subscriber and covered spouse, in the five percent sample beginning February 1970 and terminating August 1971. To determine eligibility, the subscriber had to be a member of Kaiser Health Plan beginning January 1, 1969. The only exceptions to this were spouses whose mates were members of Kaiser Health Plan January 1, 1969 and who became covered under their mate's policy after January 1, 1969. These newly covered spouses then became eligible to be interviewed. A subscriber was not interviewed, or his interview was not processed, if he was eligible January 1, 1969 but terminated Kaiser Health Plan before being interviewed (or interviewed by mistake) and did not rejoin before December 31, 1970. Of the 2,890 eligible interviews, 2,439 were processed, or 84.4 percent.[35/]

The questionnaires were developed after a review of the sociological and social psychological literature on health and medical care. Four types of variables felt to create a need for medical care were included: (1) Bio-social factors such as age and sex; (2) situational or physical environmental conditions such as type and place of work; (3) socio-cultural agents as drinking, smoking or conventional health procedures; (4) psychological stress factors, socially and environmentally produced, including financial problems and family strife. Also included were three types of variables thought to affect demand for the consumption of medical services: (1) Socio-cultural factors included the valuation placed on health, the determination of normal health states, and opinions on proper sources of medical care; (2) social components such as pressure resulting from interaction with family members, friends, and colleagues at work; (3) psychological elements such as sick role orientation, optimism or pessimism, or self-acceptance (21, p. 3).

[35/] Unpublished Kaiser paper titled: West Coast Community Surveys Household Interview Sample Statistics, October 1, 1971.

In essence, the questionnaires described information on socioeconomic-demographic characteristics and data on situational and attitudinal traits.

The questionnaires were administered separately but simultaneously to husband and wife in two parent families. A different but related questionnaire was administered to each spouse (questionnaires A and B). This approach was utilized because it was felt some data would be more reliable and valid if obtained from the husband/father and some more reliable if obtained from the wife/mother. Other data, especially that of a subjective nature such as perceptual-attitudinal data, must be obtained from each person independently (22, p. 300-301). In a one parent or single subscriber household a C questionnaire was given which contained all the questions of the A-B set. On the average it took one and one-half hours to administer the questionnaire.

Each interview was identified by the member's H. P. I. D. number at the time of the interview. All members of the subscriber's health plan were identified by unique person numbers. These numbers allowed the data from the household survey to be attached to the member's medical records.

A Comment On the Data Base

Before proceeding into a discussion of the several data bases, and their inclusion within each of the statistical models, a word is in order about the entire data base. A tremendous amount of data were collected for use in the study. Because of time and other constraints not all the data were used. To make the current presentation more concise and meaningful, only data pertaining to this part of the study will be described.

Suffice it to say that meteorological, air pollution, medical and socioeconomic data have been collected for the area bounded by Salem, Oregon on the south; Kelso-Longview, Washington and Rainer, Oregon on the nrothwest; Vancouver, Washington on the north; and Troutdale, Oregon on the east. The study area for this study has been delineated, of course, about the Portland S. M. S. A. for reasons already given.

IV. COMPILING THE MEDICAL DATA

Introduction

This chapter begins by describing the Kaiser inpatient data base, and then explains why this base was not included in this study. The second section describes the Kaiser outpatient data base, and details how this base was prepared for data processing. The final section discusses the derivation of the dependent variable for each of the statistical models, and the modifications of the variables for use in the models.

Kaiser Inpatient

A Brief Description of the Kaiser Inpatient Data Base

Since 1965 detailed socioeconomic and medical information has been obtained on each inpatient admission to Bess Kaiser Hospital. Annual inpatient admissions ran approximately nine percent of the total Kaiser membership in any given year of the study period.

Kaiser inpatient data can be broken into three major categories: socioeconomic-demographic, medical, and a functional evaluation of the patient upon discharge from the hospital. The socioeconomic-demographic data contained information on the patient's family income, education, past illnesses, type of residence (apartment, house) and state of health, among other data. Drinking and smoking habits were obtained on each adult admission. The medical data contained the patient's admitting and discharge diagnoses, date of admission and discharge, and medical services received (number of laboratory and radiology tests received). Data about the functional ability of the patient upon discharge were obtained on the patient's mobility, continence, physical rehabilitation, and activities of daily living, among other data. Functional data were not obtained if the patient was discharged to another hospital or had died.

The data collected on inpatient admissions were collectively called the inpatient data, and were recorded on a set of five magnetic tapes. Each tape was ordered by Health Plan Identification (H. P. I. D.) number and date of admission to the hospital. Three tapes contained the original inaptient and Extended Care Facility (E. C. F.)[36/] data. A fourth tape contained the records from the first three tapes which had passed a validity check for accuracy. The fourth tape also did not contain the E. C. F. data. A fifth tape contained the medical records which had not passed the validity check, and which were

[36/] An E. C. F. can be described as a unit providing skilled nursing or rehabilitation services to a patient on a less intensive basis than a regular hospital, but more intensive than a regular nursing home. At Kaiser the extended care service began April 1967 and ended December 31, 1968 when a federal grant supporting the project was not renewed. E. C. F. care now for Kaiser patients is provided by outside nursing homes.

to be eventually corrected by Kaiser.

Copies of these tapes, with the exception of the tape containing the records in error, were obtained from Kaiser Research. However, there were problems with the inpatient data as a whole which prevented its use in this study. These problems are discussed in the next section.

Reasons for Rejecting the Kaiser Inpatient Data

There were several problems and shortcomings associated with the inpatient data which made it incompatable for use with the outpatient medical data without extensive modification. For example, drugs consumed by the patient while in the hospital were not coded. The type of laboratory and radiology procedures performed were not coded in enough detail to really be of use in this study. If a patient changed H. P. I. D. numbers (because of name change, divorce, or marriage, for example) the data on the inpatient tapes were not resequenced to reflect the change; that is, the data for the patient could appear under several H. P. I. D. numbers. The main method of identifying an individual was by the H. P. I. D. number. The more numbers he had, the more difficult the situation.[37/]

[37/] One way around the multiple H. P. I. D. number problem was to use the patient's medical chart number, which was coded with the data input. Supposedly this number never changed.

The admitting and discharge diagnoses, and the surgical procedures were coded according to the International Classification of Diseases, Adapted, more commonly referred to as the I. C. D. A.[38/] Prior to January 1, 1970, the 1962 revised edition of the I. C. D. A. (43) was used for indexing inpatient records; the eighth revision I. C. D. A. (42) has been used since January 1, 1970. The I. C. D. A. numbers as between the two editions were not comparable. However, it did not appear that the eighth revision could be reconciled to the seventh revision to obtain compatability on diseases coding for the entire study period.[39/] Kaiser, thus far, has made no effort to do this.

Each surgical procedure should be related to the disease for which it was performed. Beginning in June of 1970, and not

---

[38/] The I. C. D. A. is a code used for the classification of morbidity (diseases) and mortality (death) information for statistical purposes. It is used also for the indexing of hospital records for data storage and retrieval. The I. C. D. A. are periodically revised. The 1968 edition of the I. C. D. A. has served as a basis for coding diagnostic data for the official morbidity and mortality statistics in the United States (42, p. IX-X). See (42, p. IX-XXXII) for a complete history on development and use of the I. C. D. A.

Each disease is identified by a four digit number, or has a category to which it can be referred. In addition to the disease classification there are further, supplementary classification for accidents, special conditions and examination without sickness, and surgical procedures.

[39/] Conversation with A. V. Hurtado, M. D.; Chief, Department of Medicine; Kaiser Foundation Hospitals; Portland, Oregon; January 28, 1972.

discovered until December 1971, the surgical procedures were not recorded in this manner. That is, it was impossible from the recorded data to relate the specific surgical procedures to the specific diseases. Kaiser Research was going back and recoding all the errors.

In light of the above problems, the main question to answer was whether it was worth searching the tapes for the medical records of the five percent sample members who responded to the household interview. Several preliminary analyses were made to assess the advisability of searching either set of tapes for extended care-hospital inpatient medical utilization, or just hospital inpatient utilization. The analyses made use of Kaiser statistics on hospital and extended care discharges by I. C. D. A. category (diseases of the respiratory and circulatory systems, for example) of the entire Kaiser membership for 1967 and 1968. Results from these statistics were adjusted and inferred to the five percent sample. In essence, it was found that the extra information obtained on extended care-hospital discharges, or just hospital discharges, would not justify the programming and computer expense to search 65,000 records in order to procure it.[40/] This was especially true in light of the

---

[40/] The number of respondents to the household interview, including dependents, was assumed to be 5,000 individuals. A percentage of extended care-hospital discharges per disease category to total extended care-hospital discharges was calculated for each fact the data would have had to be extensively revised and manipulated to make them compatible for use with the outpatient data.

Kaiser Outpatient

An Abstraction of the Kaiser Outpatient Medical Data

All outpatient medical care utilization by the five percent sample were recorded by medical record technicians on forms specifically prepared for computer processing. All contacts (including telephone calls and letters) with the Kaiser Health Plan system have been continuously recorded on an ongoing basis, on magnetic tape, since January 1967. Data recorded for each contact include, among others, date, time, place, and type of service. The major symptoms and duration of these symptoms were recorded for each

---

disease category. The percentages were then applied to the 5,000 sample membership to estimate how many sample members would be discharged from extended care or the hospital for certain disease categories. For 1967 hospital-extended care discharges were estimated, respectively, for the following disease categories: diseases of the respiratory system 39 and two; diseases of the circulatory system 43 and seven; neoplasms (benign, malignant and unspecified) 36 and five. In 1968, when the extended care facility was in full operation, the discharges were as follows for hospital and extended care, respectively: diseases of the circulatory system 42 and 17; diseases of the respiratory system 35 and six; neoplasms 32 and nine.

presenting and associated morbidity.[41/] The medical procedures rendered, including laboratory and X-ray, were coded in detail for each visit.

With one exception, the presenting and associated morbidities were coded according to the 1962 revised edition I. C. D. A. (43) as supplemented and adapted by Kaiser. The exception was the use of symptom codes (called "T codes") developed by Kaiser to be used in many of the cases where a disease condition was not diagnosed before patient care terminates. That is, in routine, nonhospital cases, many conditions may not be definitely diagnosed before physician care of the patient terminates. Symptoms represented the final diagnosis for ten percent of the patients utilizing outpatient services at Kaiser (28, p. 249). Finding symptoms codes in the I. C. D. A. not extensive, nor unique enough for each disease state in many instances, and to facilitate ease of coding and analyzing presenting and associated morbidity data, Kaiser developed its own

---

[41/] In the event care was given two or more morbidities during the same contact with the system, the morbidity that provoked the contact was identified as the <u>presenting</u> morbidity and the others as <u>associated</u> morbidities. The use of the term associated did not mean that the diseases were necessarily clinically linked to each other. The term means they were associated in time only for the same contact as the presenting morbidity. Nor should the refinement between presenting and associated morbidities be linked with the seriousness of either disease; a presenting morbidity could be minor, whereas the associated morbidity could be a major disease (32, p. 16-17).

symptom classification system. Each organ system was assigned a single block of numbers, and additional number groups were given to nonspecific and psychiatric symptoms. Since the I. C. D. A. have some symptom codes, Kaiser used the letter T to distinguish its codes from the I. C. D. A.

The coding system used to identify each medical procedure rendered was an adaptation by Kaiser of the 1964 California Relative Value Fee Schedule (5). Each medical procedure performed on the patient was translated into a unique four digit number for coding purposes.

The medical data system was set up in such a way that all medical services rendered could be identified with the disease for which they were performed. This was accomplished by an eight place number composed of the date upon which the service was performed, the number of visits for that day, and the morbidity number.

Many of the problems which existed with the inpatient data did not occur with the outpatient system. The same I. C. D. A. edition (43) was used throughout the four year period. The outpatient tapes were in H. P. I. D. order by date of service. The master tapes were resequenced in December 1971. All medical records appeared under the latest H. P. I. D. number as of the date of resequencing. The medical procedures rendered were coded in greater specificity thanks to the expanded use of the California Relative Value method of coding medical procedures.

However, there were two major shortcomings which existed with the outpatient data. As with the inpatient data, the consumption of drugs were not coded for the five percent sample during the study period. If a patient received medical services outside the Kaiser system such services were not included in the Kaiser data.

<u>Preparing the Kaiser Outpatient Data for Use in the Statistical Models</u>

The outpatient medical data were received from Kaiser on two magnetic types. The tapes were ordered by H. P. I. D. number and date of service. The tapes contained 145, 000 separate outpatient medical records on the five percent sample for 1967-1970. These tapes were searched by H. P. I. D. number to obtain only medical records for which socioeconomic data were available from the household interview. A new tape containing 103, 000 records was generated. The medical and socioeconomic data would be matched together by H. P. I. D. numbers.[42/]

A thorough search was made of the I. C. D. A. and T-Codes as

[42/] An extra precaution had to be taken within the program for H. P. I. D. numbers that had changed since the household interview date. The problem arose where only one member of the family unit who was interviewed changed H. P. I. D. numbers. It was highly probable with such changes that covered dependents stayed under the old H. P. I. D. number. Hence, in creating the new medical tapes, medical records were accepted under both H. P. I. D. numbers.

used by Kaiser (29) to prepare a list of over 500 diseases and symptoms which were thought to be aggrevated by or associated with air pollution.[43/] The codes were grouped into eight categories as follows: diseases of the respiratory system, allergies affecting respiratory system, other allergies and skin diseases, diseases of the circulatory system, diseases of the digestive system, diseases of the eye, diseases of the genitourinary system, and other diseases. Within the system acute and chronic diseases were grouped separately, as were benign and malignant neoplasms.

An outline of the system, including major categorical breakdowns, may be seen in Table 1. Each disease is a continuation in the specificity of the outline and, as such, is identified by a code such that any single disease or group of diseases can be analyzed. That is, each disease has been grouped together within the eight classifications on an outline basis (I-A-1-a is a malignant neoplasm of the nose, for example). The outline code (I-A-1-a) and the I. C. D. A. and T-code number were coded for each disease. Specifying I-A-1-a would allow only malignant neoplasms of the nose to be included in the analysis; specifying I-A-B would allow all diseases of the upper respiratory system to be included.

[43/] Appreciation is expressed to Sheldon Wagner, M. D.; Head, Oregon State University Environmental Health Unit; Good Samaritan Hospital; Corvallis, Oregon, for his help in developing this list.

Table 1. The disease classification system.

- I. Respiratory System
  - A. Upper respiratory neoplasms
  - B. Upper respiratory infections
  - C. Lower respiratory infections
  - D. Lower respiratory neoplasms
- II. Allergies Affecting Respiratory System
  - A. Asthma
  - B. Hay fever
  - C. Specific diseases
- III. Other Allergies and Skin Diseases
  - A. Eczema
  - B. Other allergies
  - C. Other skin diseases
  - D. Specific diseases - allergies
- IV. Diseases of the Circulatory System
  - A. Heart
  - B. Circulatory system
- V. Diseases of the Digestive System
  - A. Upper gastrointestinal tract - neoplasms
  - B. Ulcer of the upper gastrointenstinal tract
  - C. Lower gastrointestinal tract - neoplasms
- VI. Diseases of the Eye
  - A. Inflamatory diseases of the eye
  - B. Non-specific eye complaints
- VII. Diseases of the Genitourinary System
  - A. Benign neoplasms
  - B. Malignant neoplasms
  - C. Nephritis and nephrosis
- VIII. Other Diseases
  - A. Symptoms and diagnoses referrable to nervous system and special senses
  - B. Emotional, mental, and psychotic disorders
  - C. Other specific diseases

This listing was used to search the reduced outpatient medical tape discussed above, by I. C. D. A. and T-code number, to create a third outpatient tape containing only the medical records of household interview respondents and covered dependents afflicted by the diseases on the list. The "Outpatient I. C. D. A., T-Code Tape," as it is called, contained 24,000 records for the 1967-1970 period. It became the main medical tape, and was employed for generating other tapes used in the statistical analysis.

The Statistical Models: A Derivation of the Dependent Variables

Model 2.3

The California Relative Value Schedule Units. In Chapter II an index of medical services was described which would eliminate a disadvantage inherent in the use of dollar expenditures for quantifying medical services consumed. The index would have the prime advantage of dollar values or expenditures by allowing aggregation of different kinds of medical services. The index would also reflect the intensity of medical services used to treat the disease state.

No index was found which perfectly met the above criteria. However, the California Relative Value Schedule (C. R. V. S.) appeared to be an appropriate compromise and was already in extensive use by Kaiser Research for coding medical procedures. The C. R. V. S. was formulated in its first edition in 1956 by the Committee on Fees, California Medical Association. The main objective of the Committee was to develop a set of principles to govern the development of fee schedules in California, and in so doing provide some uniformity in the setting of fees where none existed before (5, p. 64). As part of this objective, the Committee also established a uniform nomenclature of medical procedures, and a standardized code to designate each procedure.

The Committee found, through a survey of the California Medical Association membership, that, while the dollar value of fees varied widely throughout the state, the relationship between fees for the same procedure remained essentially the same (5, p. 64). For example, if procedure A cost twice as much as procedure B in community X, it was found in community Y that A also cost twice as much as B, even though the dollar fees charged for each procedure could be widely different as between X and Y. Expressed as relative value units for each procedure, the relationship between A and B could adequately be reflected for the whole state in one figure. The differences in fees charged per geographical area would be taken account of in the conversion factor, taken times each unit, which changes the relative value units to dollar value.

The Committee recognized that for the C. R. V. S. to be viable it would have to be periodically revised to keep abreast of changes in the medical field.

> New methods of doing the same procedure increase or decrease the amount of time or skill required, with a resulting change in the compensation the physician should receive for the services (5, p. 65).

The Committee has put out a new edition of the C.R.V.S. every three to five years reflecting new factors of medical practice.

The C.R.V.S. was selected as the best approximation available for the desired index. For this study, its use had the following advantages. (1) All services are identified by a four digit code, and most services are weighted by a unit value. (2) These codes and unit values are applicable to an entire range of medical services: medicine, anesthesia, radiology, surgery, and laboratory for the 1964 edition C.R.V.S. (5). (3) Kaiser Research used the C.R.V.S. procedure codes, 1964 edition, to code medical services for out-patient care. (4) Kaiser, with some modification to fit their own purposes, has coded the procedure numbers and unit values for the 1964 edition C.R.V.S. on magnetic tape. The tape was continually updated by Kaiser to reflect changes in providing medical services. A copy of this tape was obtained from Kaiser. (5) The C.R.V.S. units represented, to the extent each service is accurately defined, periodically evaluated, and consistently applied (which appears to be the case with Kaiser), the best approximation of the amount of output actually provided in rendering medical services to the patient.

(6) The units can be added to quantify all medical services in one common unit, or converted into dollar value and still have an accurate approximation of the quantity of medical services consumed.

The C.R.V.S. tape received from Kaiser was modified for anesthesia procedures used in surgery. Kaiser had four anesthesia procedures used in outpatient treatment for which unit values had not been assigned. Calculation of anesthesia units in the 1964 edition C.R.V.S. is determined by adding the listed basic unit for anesthesia per surgical procedure to the time, converted to units, spent in administering the anesthesia. Kaiser had recorded on its tape only the units applicable to each surgical procedure, and not the basic units applicable to the anesthesia. The C.R.V.S. tape was modified by attaching the basic anesthesia units to each surgical procedure. Hence, whenever an outpatient surgical procedure required anesthesia, the basic anesthesia units were added to the units for the surgical procedure. Efforts to find some mean or median value of time spent on each surgical procedure was not successful. Hence, the time factor was not taken into account in calculating anesthesia units per surgical procedure.

The Index of Medical Services (Y_{ijk}). To express the effect of air pollution on health as an economic cost required that the C.R.V.S. units per medical procedure be converted to dollar values. This had to be done in a manner that would not distort the quantity of medical services consumed as measured by the C. R. V. S. units. That is, the relative value between procedures in the same medical category (surgery, for example) should not be allowed to change. This was accomplished by taking each C. R. V. S. unit per medical category times a constant dollar equivalency (conversion) provided by Kaiser. The dollar equivalences per medical category are shown in Table 2.

Table 2. Outpatient California relative value schedule dollar equivalency fees.

Medical Category	Dollar Equivalency
Surgery	$ 6.50
Pediatrics	$ 7.00
Laboratory and radiology	$ 7.00
All other	$ 8.00

The dollar equivalences came from two sources. The credit office, Bess Kaiser Hospital, provided the dollar conversions for surgery, pediatrics, and all other. The laboratory and radiology conversion factor was obtained from Kaiser Research Center. The credit office used the 1964 edition C. R. V. S., and the dollar equivalencies shown in Table 2, for billing all medical procedures to nonmembers on a fee for service basis, except laboratory and radiology services, where the 1969 edition C. R. V. S. was used (6).

The dollar equivalencies were established January 1, 1969. The equivalencies, taken times the C. R. V. S. units, reflected the going fee for services rate of professional outpatient medical services in Portland.[44/] The resulting dollar values in no way reflected Kaiser's costs of providing the services. Unfortunately Kaiser did not have a detailed enough cost accounting system to provide such information.

Laboratory and radiology procedures in the 1969 C. R. V. S. were extensively revised from the 1964 C. R. V. S. The two editions were not comparable. Hence, a dollar equivalency for laboratory and radiology services was obtained from Kaiser Research Center. This dollar equivalency was used by Kaiser to bill their Office of Economic Opportunity grant for certain laboratory and radiology procedures not paid for by Oregon State Welfare.[45/] This dollar equivalency, set in January 1969, was also held to be representative of fee for service charges for laboratory and radiology procedures in the Portland area.

[44/] Conversation with Jack Thomas, Credit Manager; Bess Kaiser Hospital; Portland, Oregon, July 13, 1972.

[45/] Kaiser was awarded an Office of Economic Opportunity grant to provide prepaid medical services to a select group of indigent families in Portland. Some of these same families were also eligible to receive medical services from Kaiser under state welfare. State welfare would reimburse Kaiser for the services provided on a set fee schedule. However, state welfare would not pay for certain procedures, which were billed by Kaiser Research to the Office of Economic Opportunity grant. (According to Henry Lamb; Financial Division; Kaiser Foundation Hospitals; July 28, 1972.)

The C.R.V.S. units were converted to dollar values (Y_{ijk}) as follows. By code number, each medical procedure on the Kaiser outpatient tape was matched against the same procedure on the C.R.V.S. tape. The units on the C.R.V.S. tape were then multiplied by the appropriate dollar equivalencies. For pediatrics and surgical procedures the appropriate dollar equivalency to assign was determined by the numeric codes assigned the doctors. The laboratory and radiology procedures were grouped on the outpatient tape within a given field, hence procedures in these fields were matched and converted to dollar values. All other procedures and respective units were taken times the eight dollar equivalency.

Certain coded medical procedures in the 1964 C.R.V.S. edition did not have unit values assigned to them. This primarily seemed to occur when the services rendered were unusual and difficult to quantify. Unless Kaiser modified the C.R.V.S. and assigned units to these procedures, no attempt was made to proxy a value for them. The problem was not widespread enough to be considered that important. However, to the extent these services were used by outpatients, Y_{ijk} would represent an understatement of the dollar value of medical services consumed.

In essence then, Y_{ijk} represents the summation of dollar values, converted C.R.V.S. units, for all outpatient medical services rendered in treating a presenting and/or associated morbidity in the same office visit per individual per day.

Model 2.5

Per Capita Costs of Outpatient Visits (Y_{icj}). The dependent variable of (2.5) represents the summation of dollar values, converted C.R.V.S. units, (expressed on a per capita basis), for all outpatient medical services rendered to Kaiser questionnaire respondents, per census tract, in treating a presenting and/or associated morbidity in the same office visit per day.

V. COMPILING THE SOCIOECONOMIC - DEMOGRAPHIC DATA

Introduction

This chapter begins by discussing the source of the socioeconomic-demographic data used in this study. The preparation of the data base for use in the statistical models is briefly summarized. The next section presents the derivation of the socioeconomic-demographic variables from the data base. The transformations and other modifications performed on the variables to make them more amenable for use in the models are detailed. The chapter ends by discussing the reasons for not including certain variables, suggested in Chapter II, within the final specifications of the statistical models.

The Socioeconomic - Demographic Data: Kaiser Household Interview

The socioeconomic-demographic data used in this study were obtained from the household interview administered to eligible members of the five percent sample. A request was submitted to Kaiser Research asking for certain data from the interview. The data were received from Kaiser on magnetic tape, with each interview ordered by Health Plan Identification (H. P. I. D.) number.

Several errors were discovered on the tape. These included

H. P. I. D. numbers that were in error, coding mistakes on several questions, and two missing interviews. A new tape was generated on which the errors were corrected, and new H. P. I. D. numbers inserted for respondents who had changed their numbers since the interview date. The latter was extremely important since the socioeconomic-demographic data from the questionnaire would be matched to the medical records by H. P. I. D. number per disease category (see Table 1).

It was decided to use only the medical records of respondents to the questionnaire for analysis purposes. The reason is that more complete socioeconomic data existed for the respondents. Many of the variables expressed for inclusion in the statistical models did not exist for Kaiser members who were dependents, or lived with the respondents to the questionnaire.

The Statistical Models: A Derivation of the Independent Socioeconomic-Demographic Variables

This section will discuss the derivation of each independent variable specified with the statistical models. Each variable in (2. 3) is also represented in (2. 5); i. e., the variables are the same, except for specification as individual or aggregate variables, respectively. Except where noted, each variable has the same hypothesized relationship with the dependent variable. Hence, the variables of the two models will be discussed simultaneously.

Most of the questions on the Kaiser household interview were coded with a numeric code, with each number representing a verbal response to the question. To alleviate the problem of having to use multiple dummy variables to segregate the answers for the same question into different response categories, the data were transformed (respecified) into continuous data wherever possible. These transformations and other modifications performed on each variable will be discussed in detail.

Age of the Patient: X_{1ijk} - Model 2.3, $\bar{X}_{1icj}$ - Model 2.5

It is expected that a higher consumption of medical services would be associated with older patients. Hence, higher values of this variable are expected to be positively related to the consumption of medical services. The age of the patient is expressed in years as of the date of medical service.

Sex of the Patient: X_{2ijk} - Model 2.3, $\bar{X}_{2icj}$ - Model 2.5

This variable is expressed as a zero if the patient is male, or a one if the patient is female in (2.3). In (2.5) the variable is expressed as the percentage of female patients seeking medical care.

The sign of the relationship between the independent variable (sex) and the dependent variable in each model will be determined by what diseases are analyzed within the dependent variables. For example, the incidence of emphysema has been reported to be ten times greater in men than women (15, p. 13). If the dependent variables were limited to include only emphysema, then one could expect the men to be more acutely ill and to consume more medical services. The sex variable, since a one represents females, would be expected to have a negative relationship to the dependent variable in each model; i.e., women consume fewer medical services than men.

The sign of the sex variable is harder to determine when many diseases are analyzed at the same time. For example, if many diseases were aggregated in the dependent variable, and the incidence of the disease, in the aggregate, per sex were not known *a priori*, then the sign of the variable cannot be predicted. That is, there may be some differences in the consumption of medical services due to sex, but *a priori* the difference cannot be predicted.

Marital Status: X_{3ijk} - Model 2.3, $\overline{X}_{3icj}$ - Model 2.5

In (2.3) this variable is expressed as a zero if the patient is not married, and a one if he is. In (2.5) the variable is expressed as the percent of married patients seeking medical care per census tract. For reasons detailed in Chapter II, this variable is expected to negatively related to the consumption of medical services.

Number of People in the Patient's Household: X_{4ijk} - Model 2.3, $\overline{X}_{4icj}$ - Model 2.5

It was desired to have some kind of density measure, such as the number of persons per room, as an indication of crowding within the patient's home. Increased crowding could result in more physical and emotional strain being placed upon all household members, and also increase the possibility of initial infection and reinfection of the household members with certain respiratory diseases. This would result in the increased consumption of medical services.

A density measure could not be obtained or specified from the Kaiser data. The number of people residing in the patient's household was the best measure that could be obtained. The variable includes not only members of the immediate family, but anyone else living in the household at the time of the interview. Higher values of the variable, while not reflecting the physical separation of household members, still reflects the increased possibility of infection and reinfection of household members with certain upper respiratory diseases resulting in higher consumption of medical services.

Household Income: X_{5ijk} - Model 2.3, $\overline{X}_{5icj}$ - Model 2.5

Household income is included as a measure of the ability of the patient's household to make expenditures, in lieu of consuming medical services, to preserve and improve their state of health. Given the basis that all Kaiser members have the same opportunity to consume Kaiser medical care, it is expected that this variable will be negatively related to the consumption of medical services.

Initially it was desired to measure the patient's permanent family income in order to eliminate transitory income changes which might occur over time. A measure of permanent family income could not be obtained directly from the Kaiser data, and there appeared to be no valid way of estimating it from the available data. Therefore, an estimate of the 1969 household income, before taxes and from all sources, was substituted in its place.

The question in the household interview obtaining data on household income was stated as follows:

> Here is a list of different yearly income groups. Could you tell me which group comes closest to the total amount that all members of this household, combined, received in the year 1969 from all sources before taxes?

It should be pointed out that in most instances the income question was asked of only one respondent per household. For example, in an A-B questionnaire set for married couples, the income question appeared on the A questionnaire but not the B, where the A questionnaire was given to the male respondent and the B questionnaire to the female respondent. If B required medical service, A's answer would be assigned to her as a measure of the household income. If two C questionnaires were administered to respondents living in the same household,[46] and the responses to the household income question were different as between the two respondents, then the two responses were averaged to obtain an estimate of household income. Out of the 2,439 questionnaires, this happened approximately five times.

Table 3. The household income question and transformations on the response categories.

Response Category	Transformations
Under $2,500 per year	$ 1,250
$ 2,500 - $ 3,499	$ 3,000
$ 3,500 - $ 4,999	$ 4,250
$ 5,000 - $ 6,499	$ 5,750
$ 6,500 - $ 7,499	$ 7,000
$ 7,500 - $ 9,999	$ 8,750
$10,000 - $14,999	$12,500
$15,000 - $19,999	$17,500
Over $20,000	$32,000

[46] This happened only when husband and wife both worked, were separate subscribers to the Kaiser Health Plan, and were both members of the five percent sample.

Table 3 shows the transformations on each response category. As can be seen, the midpoint of each category was taken as the answer for the entire category. For example, if a respondent indicated his household income was between $10,000 and $14,999, then he would be assigned an income level of $12,500 for his response, if he sought medical care. The assumption behind the transformation is that actual household income of all respondents to this category would be normally distributed about the midpoint. In actuality, however, the distribution within the category responses are probably skewed, particularly within the categories at the extremes of the income levels.

A frequency distribution was calculated for each of the nine responses on the household income question. It was found that seven percent of the households had incomes over $20,000 (the last response in Table 3). To make the last response comparable to the other categories, the median income of those who earned $20,000 or over had to be estimated. This was accomplished by using census data from the Portland Standard Metropolitan Statistical Area (S. M. S. A.).

The household income question formulated by Kaiser is in essence the same as the Bureau of Census's "income of families and unrelated individuals" for which data is published (68, p. 39-191). The use of census data made it necessary to assume that the

Kaiser membership does not materially differ from the rest of the S. M. S. A. The income categories published by the Census Bureau for the "income of families and unrelated individuals" are not the same as the Kaiser response categories in Table 3. However, there was enough similarity at the upper income levels to construct a frequency polygon. The frequency polygon was used to estimate a median income of $32, 000 for those households which indicated they earned over $20, 000.

<u>Race: X_{6ijk} and X_{7ijk} - Model 2. 3, $\overline{X}_{6icj}$ - Model 2. 5</u>

In Model (2. 3) the race variables are expressed as dummy variables in the following manner:

Race	X_{6ijk}	X_{7ijk}
Negro	1	0
Other non-white	0	1
White	0	0

On each variable the statistical test is whether the consumption of medical services is different for Negroes, or others (mostly Oriental), as compared to the consumption of medical services by whites.

In Model (2. 5) the Negro and other non-white variables are combined together, and expressed as the percentage of non-white patients consuming medical services. The variable permits a test on whether the consumption of medical services by non-whites is materially different from whites.

As with the sex variable, the relationship of the race variables to the consumption of medical services will depend upon what diseases are aggregated in the dependent variable. If the incidences of diseases included in the variable were race related, then one would expect a positive relationship between the variable representing that race and the consumption of medical services. If the relationship of race to disease incidence is not known, then, a priori, the sign of the variable's coefficient cannot be determined.

Physical Fitness: X_{8ijk} and X_{9ijk} - Model 2. 3

A physically fit patient usually has a better state of health, and hence would be expected to consume fewer medical services; i.e., these variables are expected to have a negative relationship to the dependent variable in (2. 3).

There was one question in the household interview which obtained information on physical fitness. It was stated to the respondent as follows:

> How much time and energy do you spend being physically fit - a great deal, some, or very little?

Dummy variables were used to describe the qualitative response as follows in (2.3)

Physical Fitness	X_{8ijk}	X_{9ijk}
Great deal of time and energy	1	0
Some time and energy	0	1
Very little time and energy	0	0

Admittedly the responses to the physical fitness question are subjective, and hence there is difficulty in standardizing the answers as between responses. However, this was the best measure of physical fitness available within the interview, and no other way could be seen to respecify the variable. And, secondly, the use of qualitative (dummy) variables does permit explaining some differences between individual responses.

Physical fitness was not included as a variable (2.5). The main reason was that only qualitative variables could be used to represent each response. Aggregation of qualitative variables (which represent individual, subjective responses) over census tracts, where variations in group behavior are being explained, presents a problem of adequately interpreting the results.

Drinking: X_{10ijk} - Model 2.3, $\bar{X}_{7icj}$ - Model 2.5

This variable expresses the number of times in the last 12 months, from the date of the interview, the patient has had the equivalent of about six drinks, a bottle of table wine, or eight cans of beer. Higher values of this variable are expected to be associated with a poor state of health, and hence require a greater consumption of medical services on the part of the patient; i.e., this variable is expected to be related positively to the dependent variables in both models.

Table 4 shows the transformations on each response category of the question. The transformations were constructed by taking the midpoint of each time period indicated in the response.

Table 4. Transformations performed on the responses to the drinking question.

Response Categories	Transformations
Every day or nearly every day	365.0
Three or four times a week	182.5
Once or twice a week	78.2
Two or three times a month	30.0
About once a month	12.0
Six to eleven times a year	8.5
One to five times a year	3.0
Never in the last 12 months	0.0

Cigarette Smoking Index: X_{11ijk} - Model 2.3, $\overline{X}_{8icj}$ - Model 2.5

The smoking characteristics of each individual patient were incorporated into one single index. The index took a value of zero if the patient did not smoke, and went to a high of 1960 for current smokers who had smoked the longest period of time and the largest number of cigarettes per day. The index also incorporated in it the smoking characteristics of former smokers including the amount they smoked, how long they smoked, and how long it had been since they stopped smoking.

Two sets of questions from the household interview were used to obtain information on the smoking habits of current and former cigarette smokers. The first question asked the respondent if he currently smoked cigarettes, and, if he did, how many cigarettes he smoked each day; and how many years had he smoked. The responses to this question are labeled as (I) in Table 5, which presents the transformations for each response category. The second question asked former smokers how many cigarettes they used to smoke, how many years they had smoked, and how long had it been since they had stopped smoking. The responses to this question are labeled (II) in Table 5.

In Table 5, the transformations were constructed by taking the midpoint of each response category. For the open ended responses, the following criteria were followed. A frequency distribution of the questionnaire respondents showed that fewer than nine respondents, out of 2, 439 possible, indicated they currently smoked, or had ever smoked, more than 61 cigarettes a day. It was decided that the midpoint of three packs, 70 cigarettes, would be a logical cut off point for this response category.

A frequency distribution on age was generated for the questionnaire respondents. Their median age was calculated to be 46. The legal age for smoking in Oregon is 18. Most people who do smoke have started by this age, if not before. Hence, 18 would probably be a good estimate of the median age when people start smoking. Therefore, 28 years (the difference between 46 and 18 years of age) appeared to be a reasonable cut off for those who indicated they smoked for 20 years or more.

The last response category in Table 5 (II-C), asking the former smoker how long it had been since he stopped smoking, was for over ten years. The American Cancer Society and other researchers have established that after ten years of no smoking, the death rates of former smokers and lifelong abstainers are virtually the same (3, p. 131; 82, p. 145). Given this evidence and the structuring of the last response category in (II-C), it was felt that 13. 5 years would be a reasonable cut off point for those who indicated they had

Table 5. Smoking characteristics and transformations on the response categories.

Response categories	Transformations
(I) Current smokers	
A. Number of cigarettes per day	
Under 10	5
10 to 20	15
21 to 40	30
41 to 60	50
61 and over	70
B. Number of years smoked	
Under 1 year	0.50
1 to 3 years	2.50
4 to 5 years	5.00
6 to 10 years	8.50
11 to 20 years	15.00
Over 20 years	28.00
(II) Former smokers	
A. Number cigarettes per day	
Under 10	5
10 to 20	15
21 to 40	30
41 to 60	50
61 and over	70
B. Number of years smoked	
Under 1 year	0.50
1 to 3 years	2.50
4 to 5 years	5.00
6 to 10 years	8.50
11 to 20 years	15.00
Over 20 years	28.00
C. Number of years since cessation of smoking	
Less than 3 months	0.13
4 to 6 months	0.42
7 months to 1 year	0.79
More than 1 to 3 years	2.50
4 to 5 years	5.00
6 to 10 years	8.50
Over 10 years	13.50

quit smoking over ten years ago.

There is a direct association between the number of cigarettes smoked per day, the number of years smoked, and mortality and morbidity incidences of lung cancer and coronary disease (80; 82, p. 11). It appears that as the number of years since cessation of smoking increases, the state of health of the ex-smoker becomes comparable to those who have never smoked. There is also an immediate response of the body to cessation of smoking. Within six months after a person quits smoking, the bronchial system improves to a steady state; i.e., pulmonary function or vital capacity stabilizes.[47/] There is an immediate response of the cilia to cessation of smoking. Each cigarette paralyzes the cilia[48/] for 20 minutes. It is the opinion of some medical authorities that it takes from three to six weeks after cessation of smoking to make a reasonable difference in the ability of the body to ward off infection.[49/]

It was desired to construct some form of index which would at

---

[47/] Conversation with James F. Morris, M.D.; Veterans Administration Hospital; Portland, Oregon; December 3, 1971.

[48/] Cilia are hairlike cells that line the airways and by their movement propel the dirt and germ-filled mucus out of the respiratory tract (44, p. 91).

[49/] Conversation with Arthur Koski, Head and Professor; Department of Health Education; Oregon State University; Corvallis Oregon; October 19, 1971.

least have some potential of reflecting the conditions described in the last paragraph. Extensive reading of the medical literature and conversation with medical authorities failed to give any clues as to the construction or existence of such an index. Finally, an index, described by Equation (5.1), was developed from the Kaiser data in Table 5 and used in this study.

$$S_k = \frac{(A_k + B_k) \quad (C_k + D_k)}{1 + E_k} \qquad (5.1)$$

where

S_k = smoking index of the k^{th} person seeking outpatient medical care.

A_k = the number of cigarettes per day the k^{th} patient seeking outpatient medical care <u>currently</u> smokes per day.

B_k = the number of cigarettes per day the k^{th} patient seeking outpatient medical care <u>used</u> to smoke per day.

C_k = the number of years the k^{th} patient seeking outpatient medical care <u>has smoked</u>.

D_k = the number of years the k^{th} patient seeking outpatient medical care <u>used to smoke before quitting</u>.

E_k = the number of years since the k^{th} patient seeking outpatient medical services has quit smoking; and $(k = 1, \ldots, 1)$.

If a person currently smoked 70 cigarettes per day for 28 years, his index would be 1960.0. If a patient quit smoking 13.5 years ago, and at the time he quit smoking smoked 70 cigarettes per day for 28 years, then his index would be 135.2. Patients who have never smoked in their lives would take the value zero in the index.

Higher values of the index represent more intense cigarette smoking. Because of the relationship between intense cigarette smoking and cardiovascular-respiratory problems, this index is expected to be positively related to the dependent variable in both models.

Occupational Exposure Index: X_{12ijk} - Model 2.3, X_{9icj} - Model 2.5

Job related pollution can aggrevate certain respiratory problems (31, p. 34; 45, p. 13). Job related pollution would not in most instances be measured by the regional, ambient air pollution stations. To take account of this fact, an occupational exposure index was constructed for each working patient seeking outpatient medical care. A value of zero was assigned to patients who did not work. It is expected that higher values of this variable will be positively related to the dependent variable in both models.

To be more specific, the occupational exposure index was constructed in the following manner. On the household interview there was a question which asked the respondent what his specific job was. The question was coded by the occupation classes published in the 1960 Census of Population, Alphabetical Index of Occupations and Industries (67). A listing of the occupations were submitted to Mr. Darryl D. Douglas[50/] who rated each classification on a scale of one to four. The rating was based on the potential of each job classification creating a dust problem which would aggravate certain cardiovascular-respiratory problems. The meaning of the ratings were as follows: four equaled a high hazard to dust exposure; three equaled an intermediate hazard; two equaled a moderate hazard; and one equaled a low hazard. Each weight, with its respective occupation code, was taken times the average number of hours, including regular overtime, the patient worked per day. The resulting product was the occupational exposure index.

[50/] Mr. Douglas is the Director, Occupational Health Section; Health Division; Department of Human Resources; State of Oregon; Portland, Oregon

Reasons for not Including Certain Socioeconomic-Demographic Variables in the Statistical Models

Several variables presented in Chapter II were not included in the final specification of the statistical models. This section details why the variables were not included.

The residential history of the patient was not included as a variable in the final specification of the models. In Chapter II it was argued that an individual's state of health would be better if he had lived most of his life in a rural area or other unpolluted environment. Two questions in the Kaiser interview dealt with this subject. One question asked where the respondent had lived most of the time while growing up (until his nineteenth birthday) - on a farm, small town, or city. The other question asked the respondent how long he had lived in the Portland-Vancouver area. The largest answer classification for this question was ten years or over. The answers to these questions were not considered specific enough for use in this study.

Since this study is concerned only with outpatient consumption of medical services, the availability of someone at home to take care of the patient was not considered important as an explanatory variable. The reasons for this were detailed in Chapter II.

The reading of medical literature was considered important in explaining contacts with the medical system per disease category

(Model 2.5). Two questions on the Kaiser interview dealt with the subject. One question asked the respondent whether there were any books in his home which could be used to obtain information about medicine or illness. If there were, what types were they, and did he read them fairly often, once in awhile, or seldom. The second question asked the respondent how often he read the health columns in the newspapers or magazines -- often, occasionally, seldom, or never.

Answers such as seldom, often, once in awhile, and occasionally are indistinct, and can mean different things to different people. It is difficult to standardize the answers, as between respondents, such that the replies will all have the same meaning. With individual observations qualitative variables can be used to represent the replies. Aggregation of qualitative variables, as was discussed with the physical fitness variable in the last section, presents a problem of adequately interpreting the results. No way could be seen to respecify the responses into a meaningful, simple variable. It was for these reasons that the reading of medical literature was not included in (2.5).

<u>Distance to the nearest clinic</u> was thought to be important in explaining the substitution of other medical services for Kaiser services, as the out-of-pocket costs of obtaining Kaiser services increased. The influence of this variable would be reflected through

the number of office visits. The further the patient lives from the clinic, the greater the out-of-pocket expense, and, *ceteris paribus*, the fewer number of visits made to the clinic.

Approximately 87 percent of the subscribers in the five percent sample are employed (22, p. 302). The regular clinic hours are from 9:00 a.m. to 5:00 p.m., Monday thru Saturday, although service can be obtained any time at Bess Kaiser Hospital. One could expect that many patients seeking outpatient medical care would go to the clinic from work rather than from place of residence. The percentage of those who did either could not be determined from available data. Given these circumstances, it was felt that including a distance variable could introduce more error into the model than what it could explain. And the interpretation of the variable would be difficult.

It was argued in Chapter II that some measure of the patient's tendency to contact the medical system should be included. Patients who contact the system at the first perceived sign of illness are referred to as *medical hypochondriacs.* Under the Kaiser system, as argued in Chapter II, hypochrondriac tendencies on the part of the patient should not influence the consumption of medical services (2.3), but could influence the number of contacts made with the system (2.5).

The Kaiser interview had one question which alluded to the hypochrondriac problem. The question asked the respondent that if he perceived an illness did he call the doctor right away, or did he wait awhile. If the respondent indicated he waited awhile, then the next question asked how long he waited - short time, long as can, or depends. The answers to the questions were subjective, and difficult to interpret as between respondents. Because the answers could not be respecified into a meaningful, single variable; and because of the problems associated with aggregating qualitative variables, the answers were not included as explanatory variables in (2.5).

VI. THE ADMATCH SYSTEM

Introduction

Chapter VII discusses the collection of air pollution and meteorological data, and details the assigning of air quality and meteorological values to each individual obtaining outpatient medical care. In order to understand how these values were assigned to each patient, it is necessary in this chapter to lay the necessary groundwork.

This chapter begins by discussing the Admatch program from which state plane coordinates,[51/] used in estimating the air pollution levels at the patient's residence and work address,[52/] were obtained. The next section details the collection and coding of the addresses

---

[51/] In order to insure the location of original land survey measurements, the United States Coast and Geologic Survey has worked out for each state a system of state plane coordinates (59, p. 29-30). The coordinates are expressed in feet. To limit the size of the grids and to keep scale variations to a minimum, each state usually has two or more overlapping zones. Each zone is covered by a single coordinate system (34, p. 345-346). All Oregon air pollution and meteorological stations are located in the Oregon North Zone; all Washington air pollution and meteorological stations are located in the Washington South Zone.

[52/] It is possible that the patient may face very different levels of air quality at his place of work and at his place of residence. Each measure of air quality assigned to that patient, as will be discussed more completely in Chapter VII, is weighted to reflect this fact.

for use in Admatch. The difficulties which had to be overcome after using the Admatch program are then presented.

A Description of the Admatch Program

A procedure was needed to assign air quality values to each household interview respondent who consumed outpatient medical services for the disease categories in Table 1. Two methods were available to achieve this goal. The first method was the use of isopleth maps of air pollution levels in the Portland area (see Figure 5, Chapter VII) to estimate the air pollution level at place of residence and place of employment for each working respondent. This approach was soon discarded because of the expense in generating the maps, and because of the 2,900 addresses which could be potentially used on any given day.

A second method which showed more promise was an Admatch computer program developed by the United States Census Bureau and local agencies. The program, in one of its forms, had the ability to provide state plane coordinates for each address. As will be discussed in Chapter VII, the state plane coordinates are the way air quality and meteorological values were assigned to each residence and job address.

More specifically, Admatch is an address matching system which primarily provides capability of geocoding computer readable records containing street addresses. Admatch represents a way of relating census and local data on an areal basis. It was designed primarily for use with two other programs, the Address Coding Guide (A. C. G.) and the Dual Independent Map Encoding (D. I. M. E.) file (69, p. 111).

Both the A. C. G. and the D. I. M. E. file provide similar information which can be used to obtain state plan coordinates. The A. C. G. gives census tract and block number for each address within the urbanized Portland Standard Metropolitan Statistical Area (S. M. S. A.). Each resulting census tract and block number, however, would have to be located on maps showing census tract and block numbers, and each point on the map digitized to obtain state plane coordinates. An expanded version of the D. I. M. E. file, called the Geographic Base File (G. B. F.), also gives census tract and block numbers, and in addition will provide state plane coordinates directly for each address in the urbanized portion of the S. M. S. A. Use of the G. B. F. meant that only addresses outside the urbanized S. M. S. A., and outside the S. M. S. A. itself, would require digitizing to obtain state plane coordinates.

At the time the use of these programs was being considered (November 1971), the G. B. F. was still being developed in Washington, D. C. It was to be available approximately January 1972. The A. C. G. was immediately available for use on the Portland S. M. S. A.

However, besides plotting and digitizing 2900 points, there were other problems which discouraged its use. The primary problem was that Washington and Clackamas counties had been recently readdressed, and the A. C. G. did not reflect the changes. Many Kaiser respondents to the household interview had residences and/or worked in these counties. The addresses of these respondents would have had to be located on maps without benefit of census tract or block numbers to aid in the plotting.

Given the availability of the G. B. F. within the near future, the decision was made to use it and not the A. C. G. The coding of the addresses were the same for both programs. Hence, if the G. B. F. did not become available, the A. C. G. could be used without having to revamp the coding procedures.

Address Coding for Use in Admatch

Each eligible subscriber and his spouse, who were members of the five percent sample, had what is referred to as a facing sheet made out for each of them. The facing sheets were used by West Coast Community Surveys, the firm which conducted the interviews for Kaiser Research, to contact each eligible member for purposes of setting a date and time for the interview. Included on the facing sheets were the subscriber's name, Health Plan Identification (H. P. I. D.) number, last known address and telephone number; the individual members covered by the subscriber's policy, and their person numbers, and birthdates; the type of interview (A-B set or C); and the interview date, among other information. All the facing sheets were Xeroxed, and the residence addresses were coded by Oregon State University personnel. Tight security was maintained over the facing sheets to prevent improper use. The subscriber's last name was made illegible on all the sheets, and the sheets were destroyed soon after the coding.

Each residence address computer card was identified by a letter code of R. The addresses were coded by H. P. I. D. number. The house number, street name, street type, town number,[53/] and zip code were coded for each residence address. Each address was checked in the zip code directory (74) to be sure that street names, directions, types, and zip codes were correct. If there was an error on any of these items, the address would not match with the program and would be rejected. It was found that many of the respondents' addresses were in error. Incorrect zip codes appeared to be the most prevalent problem.

[53/] Each town within Oregon and Washington had assigned to it a number by the Census Bureau. The numeric codes were obtained from a listing titled "1970 U.S. Census of Population and Housing Numeric Codes for Urban Places in Oregon," provided by the Bureau of Governmental Research and Service; University of Oregon; Eugene, Oregon. The same agency also provided the numeric codes for towns in Washington.

Also coded with each address was a number linking the address back to the facing sheet from which it was coded. This provided easy reference in case of errors. The interview date and questionnaire type (A, B or C) were coded also. In all, approximately 1500 residence addresses were coded.

Obtaining the work address for each employed respondent of the household interview was somewhat more complicated. One question in the interview asked each working respondent the name of the business or organization for which he worked. The question was not coded by Kaiser Research, but provided the only means by which the respondents work address could be obtained. Hence, by hand, all of 2,439 questionnaires were examined and the place of work coded out by H. P. I. D. number.

Phone directories were collected for all telephone districts within the study area. The addresses of the firms were coded from the phone book in the manner described for the residence addresses, except town numbers were used predominately over zip codes to designate the address place.[54] A code of <u>W</u> was used to designate

---

[54] It was possible in using the Admatch system to sort addresses on zip code or town code. In coding the residence addresses both zip and town codes were used because they were readily available. The use of zip codes were more precise and hence, less error was likely to develop from their use in the program. However, given time constraints, the job addresses used only the town numeric code.

each punched card as a work address. When multiple addresses existed for a firm and the exact address of the respondent's place of work could not be determined, the respondent's H. P. I. D. number was converted back to his name, and employee directories, obtained from numerous sources, were used to place the respondent at a specific work address. City directories (53, 54, 55, and 56) were used to provide addresses when the addresses from the phone and employee directories were not specific enough for use in the A. C. G. or G. B. F.[55/]

Respondents were not coded if their occupation indicated they were not at a fixed address for most of the working day. These included bus drivers, policemen, and salesmen, among others. An exception to this existed for the longshoreman who worked the Portland-Vancouver docks. Unlike other occupations, the area of work for the longshoreman was known. Longshoremen residing in Portland usually worked in Portland; longshoremen residing in Vancouver usually worked in Vancouver.[56/]

[55/] For example, an address such as the Foot of North Swift Street could not be used. The reverse directory, in the various city directories, where addresses were listed by street name helped to convert many of these addresses to a proper form for coding.

[56/] This is according to Thomas Percy; Pacific Maritime Association; Portland, Oregon obtained in a conversation with him on March 8, 1972.

In Vancouver most unloading took place at one terminal. Hence, all Vancouver longshoremen were coded at the terminal's location. In Portland the longshoremen could work at any job assignment they chose along the city's seven miles of waterfront. Because these assignments could change from day to day, all Portland longshoremen were coded at an address representing the midpoint of the Portland waterfront. Relative, rather than absolute, air pollution changes within the area would be reflected by this address.

There were approximately 370 respondents, out of the 1500 coded, for which business addresses could not be found, or else the correct address could not be determined from several multiple addresses. Bardsley and Haslacher, a research marketing firm in Portland, Oregon, were given the responsibility of finding job addresses for these 375 respondents.

Finally, of the 1500 individuals which were initially coded from the questionnaire, 100 did not have a place of work address coded for them. The primary reason was that the businesses for which they worked could not be located.

Each air pollution station in the study network was identified by a station number. Each station had an address, or an address was approximated for it through the use of city directories. The station addresses were coded in the same format as the residence addresses, and were designated by the letter code S. Many of the stations in the network were already located by state plane coordinates. The coordinates were coded with the addresses on those stations for which they were available. Since the coordinates were estimated from maps, which is not the most accurate way of obtaining them, it was decided to use the coordinates provided by the G. B. F. whenever they were available.

All the meteorological stations were geocoded by latitude and longitude. These had to be converted to state plane coordinates for use with the other data. The procedure for converting the geodetic positions to state plane coordinates is outlined in (34, p. 351; 40, p. 11). Values to be used in the formula for converting each latitude and longitude position are published in special tables by the U. S. Coast and Geodetic Survey (10, 11, and 12).

Many respondents to the Kaiser household interview lived or worked outside the area covered by the G. B. F. or A. C. G. Addresses from these places would be rejected by the computer programs, and would have to be processed by hand (plotted on maps and digitized) in order to estimate the state plane coordinates. Because the exact geographical area covered by either the A. C. G. or G. B. F. was not known, maps were obtained from the Census Bureau in Portland (66). The maps had census tract numbers, block numbers, and state plane coordinates. Maps for the rural routes in Washington, Clackamas, and Clark counties were obtained from the Tri-County

Community Council, the Clackamas County Planning Department, and the Clark County Public Health Department, respecitvely. Maps for rural routes for other counties were obtained from the Tscheu Publishing Company in Portland. The Oregon State Highway Department provided maps for certain cities outside the S. M. S. A.

The Admatch System Results

After numerous delays and repeated phone calls to Washington, D. C., the G. B. F. finally became available in early June 1972. Looking back, the wisdom of waiting to use the G. B. F. instead of going ahead and using the available A. C. G. is seriously open to question. The G. B. F. program was done through the facilities of the Bureau of Governmental Research and Service in Eugene. While this study was the first to use the program, the problems encountered with the agency and the program were costly. In essence, the running of the G. B. F. program with the addresses from this study was a lesson in inefficiency and unforeseen problems.

The 2,947 addresses processed through the system took three weeks to complete. Ten days, on the outside, was thought to have been adequate. The processed tapes were received back from Eugene in a computer language that could not be read by the Oregon State University Computer Center. It took a week to straighten the problem out. The biggest problem incurred with the G. B. F. was that Vancouver had six zip code zones which were used in coding the residential addresses within the city. Whoever prepared the Vancouver portion of the G. B. F. used only one zip code for the entire city. Consequently, instead of the 80 to 85 percent expected match in addresses, only a 66 percent match was obtained. Most of the non-matches occurred in Vancouver. The correction of the non-matches will be explained shortly. The job addresses fared slightly better. They were run separately from the residence and air pollution addresses. Because the job addresses were sorted through the program on town code instead of zip code, the Vancouver zip code problem was not encountered. The job addresses achieved a 72 percent match.

Other major problems encountered with the G. B. F. included missing street segments, street names not spelled correctly, and missing state plane coordinates. Rather than waste more time with the program, and additional delays in working with the agency in Eugene, the Vancouver addresses and other problems were corrected in part through the use of a program listing all addresses in the G. B. F. For each address, the program listed the state plane coordinates and census tract number. Addresses which were non-matches, rejected by the G. B. F. program, were looked up in the program, and state plane coordinates and census tract numbers[57/] were coded separately by H. P. I. D. number. Of the 925 addresses rejected by the G. B. F. program, approximately 475 were saved in this manner. Another 375 residence and job addresses were plotted on maps, digitized, and converted to state plane coordinates. Approximately 50 addresses were later discarded because the respondents lived outside the study area. The rest of the addresses represented air pollution stations on which state plane coordinates already existed.

All of the Washington state plane coordinates were expressed in feet for the Washington South Zone. In order to estimate air pollution and meteorological values for residence and job addresses, the Washington coordinates had to be converted to the Oregon state plane system This was accomplished using a program developed by Mr. Kendall B. Wood, a consulting geometrics engineer, in Portland, Oregon.

In essence every Kaiser respondent to the household interview, who lived in the study area, had a state plant coordinate for his place of residence. To the extent it could be determined, each working respondent also had a state plane coordinate for his work address.

[57/] The census tract numbers were used to group the Kaiser respondents together in like census tracts in order to obtain the per capita outpatient medical costs per census tract (Model (2. 5)).

All air pollution and meteorological stations in the study area also were located by state plane coordinates.

VII. COMPILING THE AIR POLLUTION AND METEOROLOGICAL DATA

Introduction

This chapter begins by briefly describing the function of each agency from which air pollution data were obtained. Why suspended particulate data were selected over particle fallout data as the measure of air pollution in this study will be detailed. The form in which the air pollution data were collected from each of the control agencies is presented, and the preparation of the data for inclusion in the statistical models is explained. Finally, the collection of the meteorological data and their inclusion in the statistical models is detailed.

The Air Pollution Data

An Overview of the Control Agencies

The air pollution data used in this research came from several air pollution control agencies in Oregon and Washington. These agencies are the Columbia-Willamette Air Pollution Authority (C. W. A. P. A.), the Department of Environmental Quality (D. E. Q.), and the Southwest Air Pollution Control Authority (S. A. P. C. A.). C. W. A. P. A. and D. E. Q. are Oregon air pollution control agencies;

S. A. P. C. A. is a Washington control agency.

D. E. Q. is Oregon's chief enforcement arm in managing the environment. C. W. A. P. A. is a regional air pollution control agency whose jurisdictional area includes Multnomah, Washington, Clackamas, and Columbia Counties (see Figure 5). C. W. A. P. A., as a regional air pollution control agency, operates subject to D. E. Q.'s jurisdiction. That is, C. W. A. P. A. exercises the air pollution control functions of D. E. Q. within its jurisdictional boundaries, except for five classes of air pollution sources which D. E. Q. retains under its control. These classes include aluminum reduction plants, pulp and paper mills, nuclear power plants, and motor vehicles. Jurisdiction over forestry and agricultural operations is specifically withheld from the regional agencies by legislative action (48, p. 243).

S. A. P. C. A. is a Washington regional air pollution control agency. Its jurisdictional area includes the counties of Clark, Skamania, Lewis, Wahkiakum, and Cowlitz (see Figure 5).

All regional agencies in Oregon have their laboratory procedures approved by the Department of Environmental Quality. The state pollution control staffs of Oregon and Washington have agreed on methods of measurement, analysis, data reporting, and sample scheduling. Hence, the data are comparable between the originating agencies.

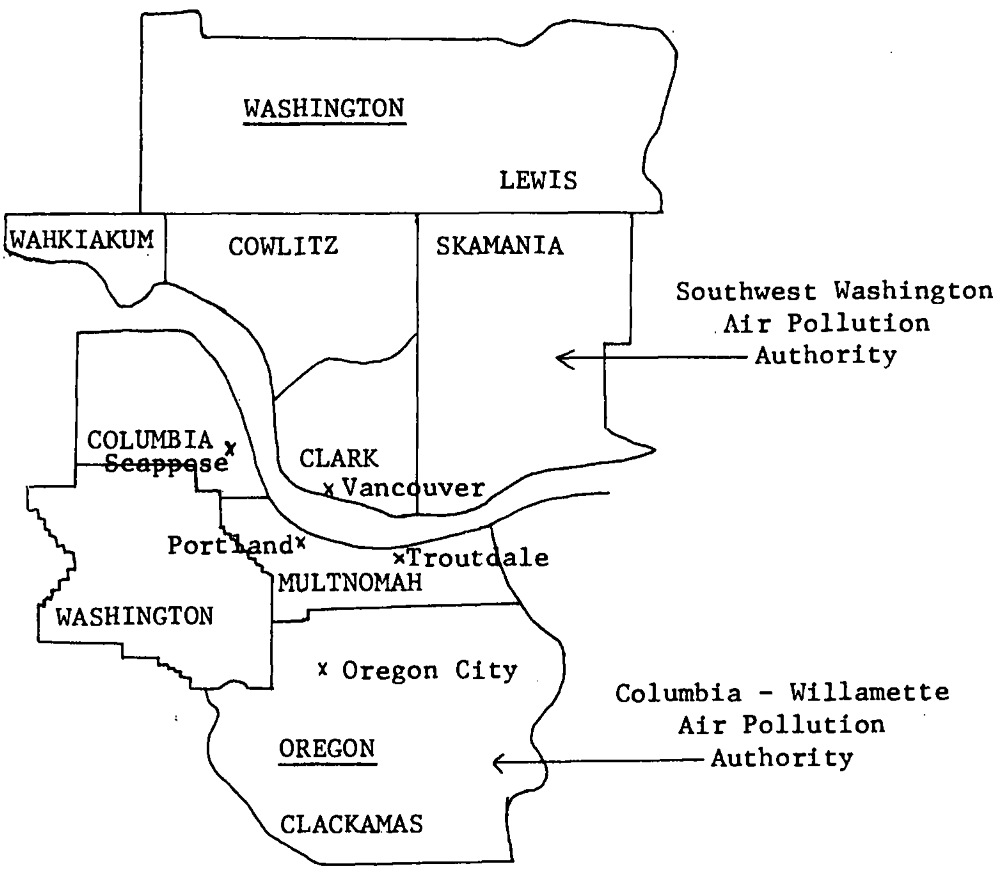

Figure 5. Boundaries of the regional air pollution control authorities.

Selecting the Proper Measure of Air Pollution

It was explained in Chapter III that Portland's main pollution problem was particulate matter. This explains the structuring of the control agencies' air pollution sampling network. The only data series of sufficient quality and quantity for the study area were measures on particulate matter: suspended particulate and particle fallout. Measures of suspended particulate were selected over particle fallout as the measurement of air quality in the study area. The reasons for this selection were many, and are presented in detail below.

The effect of particulate matter on human health were presented briefly in Chapter II. More specifically, particulate matter can physiologically affect human health in several ways. These effects would initially occur in the respiratory system, but the cardiovascular system would also be affected because of additional stress placed upon it by an impaired respiratory system. Particulate matter may exert a toxic effect via one or more of the following three mechanisms:

(1) The effect of particles on human health can be due to the particle's inherent chemical and/or physical characteristics (75, p. 141). The most important toxic aerosol is sulfur

trioxide. Others of increasing importance are lead, beryllium, and asbestos (75, p. 129).

(2) The particles may interfere with one or more of the respiratory tract's clearance mechanisms (75, p. 141). The particle size will determine the location of the toxic effect in the respiratory system. Small particles of less than two or three microns can penetrate deep into the respiratory system, where no protective mucous blanket exists (44, p. 33-34).

(3) The particle can act as a carrier of an adsorbed toxic substance (75, p. 141). The toxic effect of sulfur dioxide seems to be greater when the gas combines with an aerosol than either is alone (1, p. 62; 6, p. 111). Sulfur oxides can produce immediate airway constriction (44, p. 66).

The principal role of some air contaminants (gaseous or particulate) might be to weaken the body's defense mechanism, making it more susceptible to germs, bacteria, or viruses which can precipitate an active disease (44, p. 65). Analysis of numerous epidemiological[58/] studies indicates an association between air pollution, as measured by particulate matter accompanied by sulfur

---

[58/] Epidemiology is defined as the field of science dealing with the relationship of the various factors which determine the frequencies and distribution of an infectious process, a disease, or a physiological state in a human community (16, p. 499).

dioxide, and health effects of varying severity (76, p. 146).

High-volume samplers are used for the collection of suspended particulate matter. The operation of high-volume samplers consists of drawing air through a filter of low air resistance. The filters are felts of glass, or some synthetic organic fiber (75, p. 22). These filters can trap particle sizes as small as 0.3 microns (77). The samples are collected over a 24 hour period. The filters are then analyzed, and the results reported as total weight in micrograms per cubic meter.[59]

Particle fallout is collected in a polyethylene bucket filled with distilled water. The container is exposed at the desired location for 30 days, plus or minus two days. The contents of the bucket are analyzed and the results reported as total weight of particle fallout in grams per square meter per month.[60]

Of the two air quality measurements available for use in this study, the data on the suspended particulate matter appeared to be the more desirable. The reasons were as follows. Particle sizes

---

[59] Some agencies perform analyses on the total weight of particulates to determine their chemical composition. These analyses are also reported in micrograms per cubic meter.

[60] Chemical analyses are also performed on particle fallout, and the results reported in grams per square meter per month. The main reason for doing chemical analyses is to obtain an idea of the combustion processes and sources within the area.

collected by dust fall buckets are too large to be respirable by the human lung. Dust fall consists of particulates ten microns or larger which settle out of the air fairly rapidly (49, p. 7; 74, p. 16). Because of the settled particles large size, particle fallout measurements in no way assess the health risk of air pollution, as would suspended particulates. The large sampling period associated with particle fallout would prohibit measuring, or approximating, day to day changes in air quality. For these reasons, the total weight of suspended particulate data were used as the air pollution parameter in this study.

Collecting the Air Pollution Data

The Columbia-Willamette Air Pollution Authority. Each station in C. W. A. P. A.'s network is located on a grid coordinate system expressed in state plane coordinates. The coordinates are used by C. W. A. P. A. to plot isopleth maps of the form shown in Figure 6. The maps are designed primarily to give a pattern of pollution over the metropolitan area. Figure 6 shows the 1970, annual geometric mean isopleth for suspended particulate.

Table 6 lists, by station number, the air pollution stations from which data were taken for use in this study. Each station in C. W. A. P. A.'s network is identified by a unique six digit number. The number identifies the county, city, and site address at which

Figure 6. 1970 geometric mean suspended particulate isopleth, in micrograms per cubic meter. (Adapted from an isopleth map prepared by Columbia Willamette Air Pollution Authority).

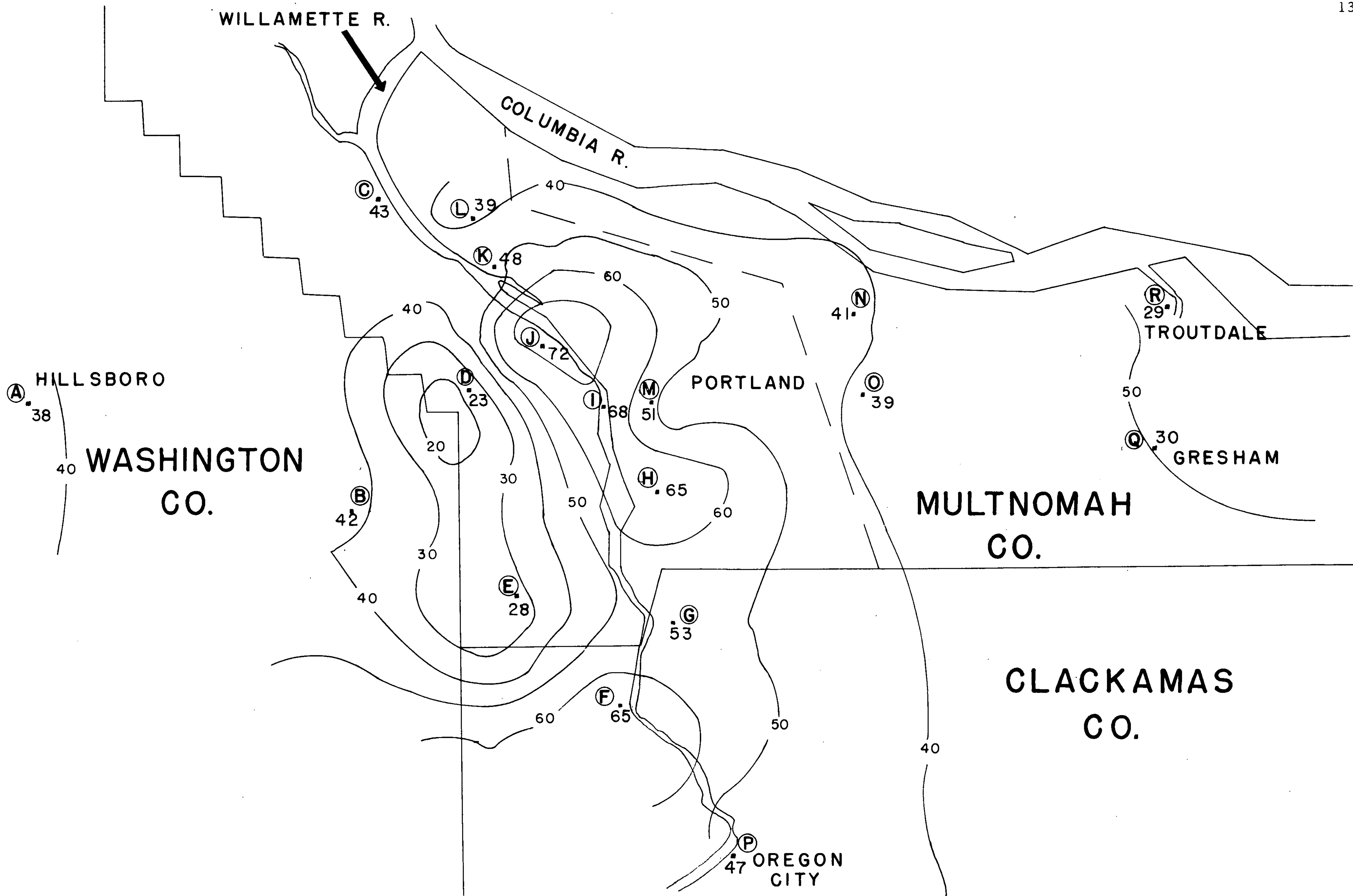
WILLAMETTE R.
COLUMBIA R.
HILLSBORO
WASHINGTON CO.
PORTLAND
MULTNOMAH CO.
CLACKAMAS CO.
TROUTDALE
GRESHAM
OREGON CITY
A 38
B 42
C 43
D 23
E 28
F 65
G 53
H 65
I 68
J 72
K 48
L 39
M 51
N 41
O 39
P 47
Q 30
R 29
20
30
40
50
60

Table 6. Air pollution stations used in this study.

Station Number	Alternative Station Number	Figure Code	Station Address	Dates of Operation	Agency
261403	-	H	1830 S. E. Schiller, Portland, Oregon	1-69 to 12-70	CWAPA
261407	-	L	6941 N. Central, Portland, Oregon	1-69 to 12-70	CWAPA
261408	-	I	55 S. W. Ash, Portland, Oregon	1-69 to 12-70	CWAPA
261410	-	E	10625 S. W. 35, Portland, Oregon	1-69 to 12-70	CWAPA
261412	-	M	1845 N. E. Couch, Portland, Oregon	1-69 to 12-70	CWAPA
261414	-	-	4906 N. E. 6, Portland, Oregon	1-69 to 6-70	CWAPA
261415	-	D	333 S. W. Skyline, Portland, Oregon	1-69 to 12-70	CWAPA
261416	-	J	3200 N. W. Yeon, Portland, Oregon	1-69 to 12-70	CWAPA
261423	-	O	340 N. E. 122, Portland, Oregon	1-69 to 12-70	CWAPA
261425	-	-	104 S. W. 5, Portland, Oregon	1-69	CWAPA
261426	-	-	1010 N. E. Couch, Portland, Oregon	4-70 to 12-70	CWAPA
261435	-	C	11212 N. W. St. Helens Rd. Portland, Oregon	4-69 to 12-70	CWAPA
261436	-	N	11717 N. E. Shaver, Portland, Oregon	4-69 to 12-70	CWAPA
261437	-	K	5000 N. Willamette, Portland, Oregon	4-69 to 12-70	CWAPA
261440	POR #77	-	3119 S. E. Holgate, Portland, Oregon	6-69 to 12-70	CWAPA
261452	-	-	14141 N. Rivergate, Portland, Oregon	6-70 to 12-70	CWAPA

Table 6. Continued.

Station Number	Alternative Station Number	Figure Code	Station Address	Dates of Operation	Agency
261431	POR #76	-	718 W. Burnside, Portland, Oregon	7-69 to 12-70	CWAPA
260801	-	Q	1300 N. Main, Gresham, Oregon	7-69 to 12-70	CWAPA
261701	-	R	GSA Warehouse-Airport, Troutdale Oregon	2-70 to 12-70	CWAPA
035501	-	P	8th and Main, Oregon City, Oregon	1-69 to 12-70	CWAPA
035504	- -	-	4th and Central, Oregon City, Oregon	4-70 to 12-70	CWAPA
035201	-	-	S. E. Cedar and Oak, Oak Grove, Oregon	12-70	CWAPA
034004	LOS #1	F	368 S. State, Lake Oswego, Oregon	6-69 to 12-70	CWAPA
034303	MIL #11	G	1550 23, Milwaukie, Oregon	6-69 to 12-70	CWAPA
036401	-	-	Fire Station, Sandy, Oregon	2-70 to 12-70	CWAPA
341001	BVN #1	B	4950 S. W. Hall, Beaverton, Oregon	11-69 to 12-70	CWAPA
343401	-	A	150 N. 3, Hillsboro, Oregon	10-69 to 12-70	CWAPA
068802	-	-	2300 W. 16, Vancouver, Washington	7-69 to 12-70	SAPCA
068801	-	-	500 W. 12, Vancouver, Washington	1-69 to 12-70	SAPCA
053101	-	-	N. W. Beacon Airport, Scappoose, Oregon	10-70 to 12-70	CWAPA
053102	-	-	100 Columbia, Scappoose, Oregon	1-70 to 9-70	CWAPA

the station is located.[61/] The station numbers were used in this study to identify each station for purposes of entering or deleting it from the study network. The reasons for doing this will be discussed later in this chapter.

All of C. W. A. P. A.'s data were kept on punched cards. For this study, all of C. W. A. P. A.'s suspended particulate data in the Portland S. M. S. A. were used. During 1969 and 1970 the suspended particulate stations of all the agencies operated 24 hours every fourth day. All the stations were operated simultaneously.

A complete listing was made of all air pollution data received from C. W. A. P. A. for the 1969-1970 study period. A table was constructed, and the months for which data were available were noted for each station. In order to determine if there were missing data, the table was compared to C. W. A. P. A.'s listing of operating stations. This list included the types of data available and the sampling period for each station. Missing stations and periods of data, to the extent they were available from C. W. A. P. A.'s records, were collected, coded, punched on data cards, and included in the study's air pollution network. Several D. E. Q. stations, for which C. W. A. P. A. had data, were identified by this process. For purposes

[61/] Each station in Figure 6 is identified by a letter code. The corresponding letter code is shown opposite the station it refers to in Table 6.

of analysis, all D. E. Q. data had to be coded and punched in the data card format that C. W. A. P. A. used. Finding the data already in the correct form eliminated having to process some of the D. E. Q. stations.

The Department of Environmental Quality. D. E. Q. had several suspended particulate stations in the S. M. S. A. which were started in the summer and fall of 1969. The operation of most these stations was taken over by C. W. A. P. A. in early 1970. These stations were referred to as joint stations. C. W. A. P. A. would collect the samples and D. E. Q. would do the analysis. D. E. Q. would then furnish C. W. A. P. A. a copy of the sample results.

Some of the D. E. Q. stations were identified by a letter code, number combination. The letters identified the town in which the stations were located; the numbers identified each station within the town. For example, with reference to Table 6 and the column labeled "Alternative Station Number" (where D. E. Q.'s numbers for the joint stations are listed), POR #76 stands for Portland station number 76; LOS #1 stands for Lake Oswego number one. The joint stations' primary identification numbers, the numbers by which they are known in this study, are those assigned to them by C. W. A. P. A. For example, station number 261440 in C. W. A. P. A.'s network is D. E. Q.'s POR #77.

D. E. Q.'s data, with the exception of the joint stations, were not available on 80 column cards. Each sample result was recorded on a data sheet. The data sheets were punched on cards, listed, and compared to C.W.A.P.A.'s listing of operating stations discussed above. Duplicate data between the D.E.Q. and C.W.A.P.A. sources were noted and eliminated.[62/] A program was written to convert the D.E.Q. card decks to the C.W.A.P.A. format.

D.E.Q. did not have state plane coordinates on any of their stations. The coordinates were provided by C.W.A.P.A. for the joint stations, or site addresses were utilized to obtain the coordinates from Admatch.

Southwest Air Pollution Control Authority. S.A.P.C.A. had two suspended particulate stations in Vancouver which were used in the study. The data from these stations were obtained from C.W.A.P.A. on cards which were in the correct format. Each station was identified by a number. The stations' site locations were identified by Universal Transverse Mercator (U.T.M.) grid numbers instead of state plane coordinates.[63/] The U.T.M. coordinates were converted to state plane coordinates.

[62/] Duplicate data existed on other stations besides the joint stations.

[63/] In large scale military mapping, the United States has adopted projection scales called the transverse Mercator and the polar sterographic. There is superimposed on the maps a rectangular grid designated the Universal Transverse Mercator (UTM) or the Universal Polar Sterographic grid (UPS). The Universal Transverse

Preparing the Air Pollution Data for Use in the Statistical Models

Use of the Spline Curve for Estimating Missing Air Pollution Data. The suspended particulate stations were operated every fourth day during the study period. The medical and meteorological data were daily observations. It was desired to observe changes in the consumption of medical services with changes in air quality. This required that a method be formulated to estimate air quality on days for which no data were available. Two alternatives were considered to estimate the data: linear interpolation and spline fit.

After considering the air pollution data base, the spline fit was chosen over linear interpolation for estimating missing air pollution observations for each station. The spline fit is accomplished by connecting each pair of adjacent points (observations) with a third degree polynomial, matching up the sections such that the first and second derivatives are continuous at each point (52, p. 404-405); i.e., the spline fit is a piecewise cubic function which is twice

Mercator is employed between 80 degrees North and South latitude; the Universal Polar Stereographic poleward of 80 degrees (59, p. 28). See (34, p. 351-355) for a more complete description of the system.

continuously differentiable (83, p. 27).[64/]

In essence, the spline curve as compared to linear interpolation, represents a higher order fit of data points. The curve is of piecewise construction, which means that only part of the data for a given station is fitted at a time. The spline fit, as employed in this study, used four observation points to fit each segment. For example, observation points A, B, C, and D would be used to fit one segment of the curve. Points B, C, D, and E would be used to fit the next segment. While estimated points between D and E would be most influenced by the actual observed values at D and E, points B and C would also have some minor influence. Linear interpolation uses only two observation points to determine each linear segment (D and E, for example).

There is an advantage of having other observation points, besides the two closest in time, playing some role in estimating missing observations. The first and second derivatives indicate the average rate of change and rate of change, respectively, at the observation points. The signs of the derivatives reflect the direction of change. Hence, if there is an inversion which causes pollution to increase, then to some extent, the inversion will be reflected in the curve and the estimated values.

[64/] For those who are interested in a strict mathematical treatment of the spline fit the following sources are recommended (52, p. 404-461; 83, p. 27-31).

With linear interpolation this would not be the case for the following reason. Only two observation points determine the construction of each linear segment, and the rate of change along each segment is assumed constant, which, in reality, may not be the case.

In essence then, either method, spline fit or linear interpolation, could have been used to estimate missing values. While neither fit gives specific information as to what is actually happening between any of the four day observation points, it was felt that the spline fit would be a more accurate representation of real world conditions.

Assigning Air Pollution Values to Residential and Job Addresses. As mentioned in the previous chapter, a procedure was formulated to assign air quality data to each patient receiving outpatient medical care from Kaiser. Two techniques were considered to estimate the values: least squares analysis and the use of the three closest stations to calculate a weighted value of air pollution.

The least squares polynomial was the first model tried. Twenty air pollution stations, on which observations existed for a given day and for which state plane coordinates were known, were used to test the feasibility of the least squares method. Each station's observations were regressed against the station's state plane coordinates.

The least squares model was an attempt to estimate (represent) air pollution values on a geographic surface using X and Y coordinates.

This is similar, in method, to the isopleth map illustrated by Figure 6. However, the computer program used to draw the isopleth map was a more complex one than the least squares models tested in this study. The isopleth program interpolated between data points to fit a geographic surface with constant pollution values.[65] With least squares, X and Y coordinates were used to represent pollution values at various points. The model assumed some relationship between air pollution and the state plane coordinates. The model was then tested to see how well it fit the data. These tests are now briefly examined. The general form of the model may be seen in (7.1).

$$P_{ij} = f(X_{ij}, Y_{ij}) \qquad (7.1)$$

where

P_{ij} = the air pollution values for station i on day j.

X_{ij} and Y_{ij} = the state plane coordinates, expressed in feet, for station i on day j; and

$(i = 1, \ldots, t)$; $(j = 1, \ldots, m)$.

Different least squares polynomials were tried which were

---

[65] The program used in drawing Figure 6 was a leased program from International Business Machines, Inc. Because computers from Control Data, Inc., are used at the Oregon State University Computer Center, the program could not be utilized here.

linear, quadratic, and cubic in (X_i, Y_i).[66/] The multiple coefficient of determination (R^2) for each model was 0.065, 0.305, and 0.338, respectively. While the R^2 improved going from a quadratic to a cubic function, it did not improve materially. Hence, it was felt that going to a higher order model would not be worth the computer expense. And, because of the low R^2, it was thought the least squares models would not provide any more reliable air pollution estimates than using a weighted value of air pollution.

The weighted value of air pollution, estimated for each patient's residence and job address, was calculated by weighting the observed air pollution values of the three closest stations by their distance from the patient's addresses. The concept is illustrated by (7.2).

$$P_{ij} = \frac{1/D^2_{1ij}(S_{1ij}) + 1/D^2_{2ij}(S_{2ij}) + 1/D^2_{3ij}(S_{3ij})}{1/D^2_{1ij} + 1/D^2_{2ij} + 1/D^2_{3ij}} \quad (7.2)$$

[66/] More specifically, the polynomials were of the following form: $P_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 Y_{2ij} + \epsilon$; $P_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 Y_{2ij} + \beta_{11} X^2_{1ij} + \beta_{22} Y^2_{ij} + \beta_{12} X_{1ij} Y_{2ij} + \epsilon$; $P_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 Y_{2ij} + \beta_{11} X^2_{1ij} + \beta_{12} X_{1ij} Y_{2ij} + \beta_{22} Y^2_{2ij} + \beta_{111} X^3_{1ij} + \beta_{112} X^2_{1ij} Y_{2ij} + \beta_{122} X_{1ij} Y^2_{2ij} + \beta_{222} Y^3_{2ij} + \epsilon$; where $\beta_0, \ldots, \beta_{222}$ are the parameters to be estimated; P_{ij}, X_{ij}, and Y_{ij} are as defined in (7.1) and $\epsilon \sim N(0, \sigma^2)$.

where

P_{ij} = the estimated air pollution value for address i on day j.

D_{1ij} = the distance (in feet) of the closest station from address i on day j.

S_{1ij} = the observed air pollution values of the closest air pollution station to address i on day j.

D_{2ij} = the distance (in feet) of the second closest station from address i on day j.

S_{2ij} = the observed value of air pollution on day j for the second closest air pollution station to address i.

D_{3ij} = the distance (in feet) of the third closest station from address i on day j.

S_{3ij} = the observed value of air pollution on day j for the third closest air pollution station to address i; and (i = 1, . . . , d); (j = 1, . . . , m).

The distance from each address to each station was calculated using state plane coordinates. As can be seen from (7. 2), each air pollution observation is weighted by the inverse distance squared from the address. The air pollution value observed at the nearest station to the patient's addresses receives the greatest weight.

The use of distance to weight each air pollution observation is simple and inexpensive to employ. The biggest disadvantage in its use is that only distance is taken into account. Topography and meteorological conditions can play an important role in influencing the distribution of pollutants.

Another disadvantage of using the system is that the air pollution values calculated for any address are limited to being between the highest and lowest observed pollution values of the stations used in the formula. However, given the structuring of the study's air pollution network, this disadvantage is not that critical for the following reason. Most air pollution stations used in this study are designed and located to measure air pollution in an air mass; i. e., the stations are designed to measure air pollution in the ambient air. The sampling sites, for the most part, are representative of the general surroundings, and are not contaminated by any special source. Hence, it is not unreasonable to expect the air pollution at any given address to be most generally represented by the observed air pollution values of the station closest to it.

Several limitations were placed on the system of estimating air pollution values for given addresses. It can be seen from Table 6 that certain stations located in the immediate Portland area did not operate continuously during the two year study period. These stations were 261425, 261426, 261435, 261436, 261437, 261440, 261452, 261431, 035201, and 068802. When these stations were not in operation the computer program (by being aware of the station numbers), calculating the weighted air pollution values, would account for it. If any of these stations happened to be one of the three closest stations, they would be passed over and the next closest station would be used to calculate the weighted air pollution value.

Other air pollution stations, located outside the immediate Portland area, also operated only part of the study period. A listing of the station numbers are as follows: 260801, 261701, 034004, 034303, 036401, 341001, 343401, and 053102. Until the stations became operable, any patient living in the area, where one of these stations would be the closest station to his residence address, had his medical records excluded from this study. For many of these outlying areas the next closest station would have been somewhere in Portland, which would have been too far away to have been very representative.

In using the spline fit, if the period between sample observations on any station exceeded eight days, the station would not be used in calculating the weighted air pollution values. A value of zero would be assigned to the station and the next closest station would be used in the calculation.

Derivation of the Ambient Air Pollution Variable: X_{13ijk} - Model 2.3, $\overline{X}_{11icj}$ - Model 2.5

Many of the Kaiser respondents worked, and the place of work, in all but a few instances, was different from the residence address. One would expect air pollution values on the same day to be different between the residence and work address. It was desired, therefore, to reflect these differences in the air pollution value assigned to each patient. The air pollution variables in each model were weighted to reflect the patient's potential exposure to air pollution at both residence and job address.

The weighted air pollution value assigned to each patient was calculated by (7.3).

$$\overline{P}_{ijk} = \frac{\left[(J_{ijk} + T_{ijk})\, P_{ijk}\right] + \left[24 - (J_{ijk} + T_{ijk})\right] P^{*}_{ijk}}{24} \qquad (7.3)$$

where

$\overline{P}_{ijk}$ = the weighted air pollution value assigned to the k^{th} patient who consumes outpatient medical services for disease i on day j.

J_{ijk} = the number of hours worked (including regular overtime) per day of the k^{th} patient consuming outpatient medical service for the i^{th} disease on day j.

T_{ijk} = the average transit time to and from work of the k^{th} patient consuming outpatient medical services for the i^{th} disease on day j.

P_{ijk} = estimated air pollution at the job address for the k^{th} patient consuming outpatient medical services for the i^{th} disease on day j.

P^{*}_{ijk} = estimated air pollution at the residence address for the k^{th} patient consuming outpatient medical services for the i^{th} disease on day j; and

$(i = 1, \ldots, n); j = 1, \ldots, m); (k = 1, \ldots, l).$

From (7. 3) it can be seen that a weighted mean of air pollution, based on a 24 hour day, was assigned to each patient where the weights were the number of hours on the job plus transit time to and from work, and the number of hours at place of residence. For purposes of calculating the weighted air pollution, it was always assumed that the patient was either at work, in transit, or at his place of residence. Some modifications to this scheme, to take account of weekends and other days off, were initiated. These will be discussed shortly.

From the household interview it was determined whether the patient worked or not, and the transit time to work. The transit time to and from work was estimated using a question from the household interview which asked the respondent how long it took him to get to work. The responses, and the transformations on each response, may be seen in Table 7.[67/] The times were doubled to reflect transit time from as well as to the place of work. The transit time was added to the number of hours on the job, because it was felt the pollution in transit to and from the job site would be higher than air pollution at the place of residence. If the patient did not work, the job and transit time would be zero, and the patient would be assigned the air pollution value at his place of residence.

Table 7. Transit time to the place of work.

Response category	Transformation
Under 10 minutes	5 minutes
15 to 30 minutes	20 minutes
Over 30 minutes	45 minutes

Information from a question in the household interview, which asked the respondent what days of the week he normally worked, permitted days off the job to be accounted for in (7.3). If the patient indicated he worked Monday through Friday, then Saturday and Sunday would be considered his days off. If he became ill on

[67/] Again, the midpoint of each response was used for purposes of the transformation. A 45 minute cut off was placed on the last response. Forty-five minutes is a reasonable transit time, if one is familiar with the best routes, for a patient having a job in one end of the study area and a residence at the other end.

Saturday or Sunday, and consumed outpatient medical services, the value of air quality assigned to him would be his residence air pollution. Unless it was indicated otherwise, Sunday was always considered a day off for all work patterns considered in this study. An exception to this was when the patient indicated he worked through the weekend, but had a day off in the middle of the week. In this case Wednesday was assumed to be that day. A basic assumption was that those who had Tuesdays and Thursdays off would be normally distributed about Wednesday. Patients who indicated they worked less than five days a week were assigned their residence air pollution values, because there was no way of telling, from the Kaiser data, which days they worked.

For reasons already detailed in this chapter and in Chapter II, it is expected that weighted air pollution will have a positive relationship to the dependent variable in both models. That is, higher values of air pollution will result in greater consumption of medical services.

Collecting the Meteorological Data

Characteristics Required of the Meteorological Data

The meteorological data had to meet two main criteria for use in this study. First, the meteorological station site had to be representative of a large geographic area. Secondly, the data, if taken from more than one source, had to be comparable as between sources, or else be of such a form they could be made comparable. Comparability would include the method of obtaining the data, the units in which the data were recorded, and the competency of the originating agency.

In selecting meteorological data an attempt was made to take account of the meteorological influences of the Columbia River Gorge. Conditions may be quite different at the same time between North Portland and downtown, or the rest of the Willamette Valley. For example, during the winter it is not uncommon to see it snowing in North Portland near the Columbia River, but raining in downtown Portland (or mixed rain and snow). The differences result because of the interaction of the dry, cold east gorge winds with the moist marine air from the west. Table 8 illustrates this phenomenon using using 1967-1970 annual rainfall and temperature data from these meteorological stations: Oregon City (located in the Willamette Valley), Portland Weather Bureau City (located in downtown Portland), and Portland Weather Bureau Airport (located at the Portland International Airport close to the Columbia River). In all years notice that Portland Weather Bureau Airport has less rainfall and colder temperature than Oregon City or Portland Weather Bureau City.

Table 8. Average annual precipitation and temperature at selected meteorological stations.

Year	Oregon City Precip. (inches)	Oregon City Temp. (°F)	Portland Weather Bureau City Precip. (inches)	Portland Weather Bureau City Temp. (°F)	Portland Weather Bureau Airport Precip. (inches)	Portland Weather Bureau Airport Temp. (°F)
1967	40.87	55.50	38.97	56.50	29.24	54.30
1968	65.44	54.40	59.99	55.10	50.89	52.80
1969	56.15	53.90	45.97	54.60	37.11	52.80
1970	53.23	54.40	52.64	54.90	41.35	53.60

Source: U.S. Department of Commerce. Environmental Data Service. Climatological Data, Oregon. Annual Summaries 1967 - 1970. Vols 73 to 76, no. 13.

Sources of the Meteorological Data

The National Weather Service (N. W. S.) was the only source for meteorological data that was of sufficient quality and quantity to be used in this study.[68/] N. W. S. meteorological stations can be broken into two classes. (1) First Order Stations which are located primarily at airports, and provide the most complete meteorological data (temperature, weather types, precipitation, barometric pressure, wind speed and direction, percent sunshine, percent sky cover, relative humidity, and visibility, among others). The data

[68/] In investigating and retrieving data available through the National Weather Service, the researcher is indebted to Stanley Holbrook; Oregon State Climatologist; National Weather Service; Portland, Oregon.

are presented in monthly issues of Local Climatological Data which are published by station. The First Order Station in the study area was at Portland International Airport. (2) Second Order Stations included all the other meteorological stations. The data was presented monthly in the Climatological Data by state. The data from these stations were not as numerous as they were from the First Order Stations. Primarily only daily precipitation and temperatures were consistently provided by the Second Order Stations.

Several sources (13, 14 and 46) were examined to determine whether the meteorological data were readily available on magnetic tape or punched cards. They were not, and a request was submitted to the State of Oregon Climatologist for all published Climatological Data for Washington (71) and Oregon (70), and all published Local Climatological Data for Portland International Airport (72). The meteorological data used in this study were coded from these publications.

Coding the Meteorological Data

Two problems were encountered with the meteorological data which had to be resolved before the data could be coded and punched on cards. The first problem was what stations, of those available, to include in the study's meteorological network. The second problem was the frequency with which the data should be coded, i.e., hourly, daily, or monthly.

There were many stations and data which could be coded within the study area. Several simple statistical tests and visual inspection of the data were conducted to determine what stations should be used, and the frequency with which the data should be coded. The analyses dealt with the variations of the meteorological conditions in the study area over space and time. The period of record for these tests was 1967 through 1970.[69/]

When one year was compared to another during the 1967 to 1970 period, mean annual temperature and total annual precipitation of all meteorological stations in the study area, with one exception, fluctuated together. That is, if Hillsboro was greater in mean annual temperature for 1967 compared to 1968, then so were the other stations. The exception was Oregon City on annual precipitation between 1969 and 1970.

A statistical test was made to determine whether the difference between the maximum and minimum annual temperature means were statistically different. The two stations exhibiting the largest difference between each other were Portland City, annual mean

---

[69/] The stations used in these analyses were located in or near the towns of Warren, Portland, Hillsboro, Oregon City, Troutdale, and McMinnville in Oregon; and Vancouver, Washington.

temperature 1967, versus Vancouver, annual mean temperature 1970. The test showed the means not to be statistically different at the five percent level. A statistical test of the maximum and minimum monthly temperature means, for the same months of the four year period on all stations, showed the means to be statistically different from each other at the one percent level. The two stations exhibiting the largest difference were again Portland City and Vancouver for February 1970. From these analyses it appeared that while there may be monthly fluctuations between stations, the annual meteorological conditions within the study area were not materially different with respect to temperature. Where changes did occur from one year to the next, with temperature and precipitation, the changes, with the exception noted, were all in the same direction.

Since the study was using daily medical records, and the air pollution values could be estimated daily, it was decided to use daily meteorological conditions. However, a judgement was made that, since all the stations seemed to fluctuate in the same direction, coding of daily data for all the stations would not add that much to the statistical analyses. Hence, the decision was made to code the daily observations of two stations and let them represent the meteorological changes over the study area. The stations coded were Portland International Airport, located on the Columbia River, and Portland Weather Bureau City, located at Northwest Eighth and

Davis Street in downtown Portland. The Portland Airport station was assumed to be representative of the effects caused by the Columbia Gorge. The Portland City station was selected because it was removed, somewhat, from the influence of the Gorge. Data from the downtown station would be more representative of the southern portion of the study area than data from Portland International Airport. Patients consuming medical services were assigned the meteorological values of the station closest to their residence address as determined by state plane coordinates.

Derivation of the Meteorological Variable: X_{14ijk} - Model 2.3, X_{10icj} - Model 2.5

The effect of temperature extremes on cardiovascular-respiratory problems were detailed in Chapter II. It was desired to specify a measure which would incorporate both temperature extremes- hot and cold - into one variable. Degree days, the absolute difference between the average daily temperature and 65°F, appeared to fulfill this requirement. As the average ambient temperature becomes colder the variable will increase; i.e., the absolute difference between the cooling temperatures and 65°F will increase. And as the average ambient temperature becomes warmer the variable also will increase; i.e., the absolute difference between the warming temperature and 65°F will increase. As the variable increases, one would expect increased aggravation of certain cardiovascular-respiratory problems. Higher values of this variable are expected to be positively related to the dependent variables of both statistical models.

VIII. STATISTICAL RESULTS

Introduction

The first section of this chapter will present some of the statistical analyses performed on the air pollution data from Portland. Several preliminary tests of Models (2. 3) and (2. 5) were made. For the tests a ten percent systematic sample of all available observations was drawn. As a result of the tests, several variables in the statistical models were respecified. The last part of the chapter will detail the statistical results obtained from the respecified models, where all of the data were included in the analysis.

Preliminary Analyses on the Air Pollution Data

One of the initial questions which had to be answered, before this study could be undertaken, was whether there was variation over time and space of air quality levels within the Portland area. Several preliminary analyses were conducted on the air pollution to determine the existence of such variation. The results of the analyses are discussed in the next several paragraphs.

One form of analysis included computing the monthly means, variances, and standard deviations for nine suspended particulate stations which had been in operation throughout the 1969 to 1970 study period. The results indicated that there was quite an amount of within month variation for each station. This was expected.

An analysis of variance was run on the monthly means of the nine suspended particulate stations for the 1969 to 1970 period. The analysis of variance table may be seen in Table 9. With an analysis of variance containing interaction terms, if any of the first order interaction are significant (year by month, year by station, and month by station in Table 9) then the main effects (year, month, and station in Table 9) do not have any meaning.

Table 9. Analysis of variance table for the nine suspended particulate stations.

Source	Sum of Squares	Degrees of Freedom	Mean Square	F
Year	3967.2244	1	3967.2244	-
Month	17909.9682	11	1628.1789	-
Station	102040.3072	8	12755.0384	-
Year by month (YXM)	16389.1217	11	1489.9202	3.2279
Month by station (MXS)	5398.1639	88	613.3882	1.3289
Year by station (YXS)	3881.0485	8	485.1311	1.0510
Year by month by station (YXMXS)	40618.4404	88	461.5732	-
Total	238784.2743	215	-	-

Looking at the first interaction term, (YXM) the following hypothesis is being tested: Ho_1 : The differences between months are the same from one year to the next. The null hypothesis is rejected. The F test is significant at the one percent level. The difference between identical months are not the same from one year to the next. One could tentatively conclude that the depth of the pollution surface (amount of pollution) in Portland changes from one year to the next within the same months.

The interaction MXS tests the following hypothesis: Ho_2 : The differences between stations are the same from one month to the next averaged over the years. The F test does not allow a rejection of the null hypothesis. One can tentatively conclude that the differences between stations do not change significantly from one month to the next. This would tend to indicate that while the depth of the pollution surface might change by month from one year to another, the shape of the surface will not change significantly from station to station over time.

The interaction YXS tests the following hypothesis: Ho_3 : The differences between stations are the same from one year to the next. The F statistic does not allow a rejection of the null hypothesis. This would tend to collaborate the MXS interaction: While there may be changes in the depth of the pollution, the relative shape of the surface does not change from one year to the next.

The tests of the above hypotheses allow the following conclusions to be advanced about air pollution patterns in Portland. If annual or monthly isopleths of pollution loadings in Portland were compared, one would find that while the amount of pollution (depth of the pollution surface) in Portland may vary over time, the relative loadings (shape of the pollution surface) will remain unchanged, i.e., those areas with high pollution will remain high in comparison to all the other areas, those areas with low pollution will consistently remain low in comparison to all the other areas.

Preliminary Examination of the Statistical Models

Only outpatient medical records of respondents to the household interview were examined in this study. To obtain these records a sort was made on the "Outpatient I.C.D.A., T-Code Tape," discussed in Chapter IV, to create a second magnetic tape containing only respondent outpatient medical records for 1969 to 1970. This tape will be referred to as "The C.R.V.S.[70/] - Severity Tape." It contained approximately 6,000 outpatient medical records.

The medical records were matched to the socioeconomic-demographic data obtained from the "Household Interview Tape." As the medical and socioeconomic-demographic records were

[70/] C.R.V.S. stands for California Relative Value Studies.

being matched, the records of 187 respondents who did not live in the study area were deleted.[71/] The town and zip codes used in coding the residence and job addresses for Admatch were utilized to delete the records.

Air pollution and meteorological values were estimated for each patient on the day of contact with the medical system, and for each of the three days prior to that contact. The lagging of air pollution and meteorological values is an attempt to answer how soon after meteorological and air pollution conditions change will morbidity patterns change.

There are two principle types of delays which can affect a postponement between exposure and the seeking of medical care. One, it may take a short time after exposure for a disease to become aggravated. Two, once a patient is ill, it may take him awhile to recognize it and seek medical care. In essence then, the quest for a time sequence in causality between exposure to air pollution and meteorological phenomena, and the onset of an illness, presupposes some reasonable length of time between cause and effect. Because several medical studies, using daily data, have used three day lags, (19, p. 1062-1063; 26, p. 594), this study did likewise.

[71/] These 187 respondents included the respondents living at the 50 addresses which were indicated as being outside the study area in Chapter VI.

Two statistical models were postulated for this study. The transformations which might be required on the variables within the models, in order to obtain the best fit of the data by the models, were not known beforehand. There were many transformations which could be performed on each of the variables. Using all the data to test the transformations would be expensive and time consuming.

There exists at the Oregon State University Computer Center a program titled Statistical Interactive Programming System(S. I. P. S.). S. I. P. S. provides direct and immediate interaction with the computer. This is a distinct advantage to the researcher in testing different specifications of statistical models and individual variables within models (either one at a time or in combination with each other). Another advantage in using S. I. P. S. , besides its speed, is that it is relatively inexpensive for the amount of information that can be obtained from it. The biggest disadvantage with S. I. P. S. is that there is a limit to the amount of information (specifically the number of variables and observations) that it can handle. Therefore, to take advantage of S. I. P. S. , a ten percent, systematic sample was drawn from the C. R. V. S. -Severity Tape.

The ten percent sample was used to perfect the statistical models (2. 3 and 2. 5), in terms of final specification, before using all of the data to test them. A different ten percent sample was used to perfect Model (2. 5). The reason for this was because of the aggregation over census tracts, which required the writing of a special program and a different ordering of the tape from which the sample was drawn.

Many transformations were tried with each of the variables in the models. The method used to develope each transformation was to look at scatter diagrams between the dependent and each of the independent variables. The form the scatter would take would often provide clues to the transformations needed on the variables. The transformations were then tried in regression. Where several transformations were tried on a variable in the model, the criterion used to decide between the specification was to choose the transformation giving the highest level of significance; i. e., the largest t statistic on the coefficient. Where this did not discriminate, the transformation which yielded the highest coefficient of multiple determination (R^2) was chosen. The following variables in both models were transformed to logarithms base e: the dependent variables, the air pollution variables (including all lags), and the meteorological variables (including all lags). The meteorological variable was respecified from degree days to a Temperature Humidity Index (T. H. I.).

A more complete discussion of the T. H. I. will be given in the next section. Suffice it to say that it is an empirical measure of the impact on human comfort of temperature and humidity (38, p. 56). The higher the value of the index, the greater the patient's discomfort.

Degree days was dropped from the models because the coefficients were not statistically significant, and had negative instead of the expected positive signs. The discussion about the meteorological variable in Chapter II demonstrated there are many meteorological conditions which can affect a state of health. A. priori, it is difficult to identify the meteorological conditions which will affect a state of health within a given region; i. e., different meteorological variables must be tried. Degree days was included within the models, because it was a variable which took account of extremes of temperature at both spectrums. Probably the prime reason the variable was not significant was the lack of variation in the variable. For example, on the ten percent sample, degree days had a mean of 13.65 and a standard deviation of 8.89 for the unlagged variable. This would tend to indicate that the temperature extremes in Portland were not severe enough to affect a state of health.

The respecified statistical models resulting from S. I. P. S. are given below. Model (2.3) has been respecified as Model (8.1) and Model (2.5) has been respecified as Model (8.2).

$$\ln Y_{ijk} = \ln\beta_0 + \beta_1 X_{1ijk} + \beta_2 X_{2ijk} + \beta_3 X_{3ijk} + \beta_4 X_{4ijk} + \beta_5 X_{5ijk} + \beta_6 X_{6ijk} + \beta_7 X_{7ijk} + \beta_8 X_{8ijk} + \beta_9 X_{9ijk} + \beta_{10} X_{10ijk} + \beta_{11} X_{11ijk} + \beta_{12} X_{12ijk} + \beta_{13} \ln X_{13ijk} + \beta_{14} \ln X_{14ijk} + \epsilon \quad (8.1)$$

where

X_{14ijk} = a measure of the meteorological conditions that the k^{th} patient, who consumes medical services for disease i, is exposed to on day j, expressed as a Temperature Humidity Index; and $(i = 1, \ldots, n)$; $(j = 1, \ldots, m)$; $(k = 1, \ldots, l)$.

And $X_{1ijk}, \ldots, X_{13ijk}$, $\beta_0, \ldots, \beta_{13}$, and ϵ are as previously defined with Model (2.3).

Model (2.5) is respecified as follows.

$$\ln Y_{icj} = \ln\beta_0 + \beta_1 \bar{X}_{1icj} + \beta_2 \bar{X}_{2icj} + \beta_3 \bar{X}_{3icj} + \beta_4 \bar{X}_{4icj} + \beta_5 \bar{X}_{5icj} + \beta_6 \bar{X}_{6icj} + \beta_7 \bar{X}_{7icj} + \beta_8 \bar{X}_{8icj} + \beta_9 \bar{X}_{9icj} + \beta_{10} \ln\bar{X}_{10icj} + \beta_{11} \ln\bar{X}_{11icj} + \epsilon$$

where

$\bar{X}_{10icj}$ = the average measure of meteorological conditions for patients consuming outpatient medical services for disease i from census tract c on day j, expressed as a Temperature Humidity Index; and

$(i = 1, \ldots, n)$ $(c = 1, \ldots, p)$; $(j = 1, \ldots, m)$.

And $\bar{X}_{1icj}, \ldots, \bar{X}_{9icj}$, and $\bar{X}_{11icj}$; $\beta_0, \ldots, \beta_9$, and β_{11}; and ϵ are as previously defined in Model (2. 5).

There were several items accomplished through the S. I. P. S. program which influenced the testing and respecification of the statistical models. A test was made for an interaction between the air pollution and meteorological variables, with the meteorological and air pollution variables also specified in the model at the same time. The test proved not to be statistically significant.

Two air pollution variables were tested in S. I. P. S. One variable was composed only of the patient's estimated exposure to air pollution at his place of residence. The second variable tested was weighted, as per the discussion in Chapter VII, to reflect exposure to air pollution at place of residence and work. The weighted air pollution variable proved to be statistically more significant.

It was felt that if a relationship between air pollution and the consumption of medical services were to exist, that association would most likely be with respiratory and circulatory illnesses (categories I and IV in Table 1, Chapter IV). Air pollution was found to be statistically significant at the five percent level for respiratory diseases analyzed alone, and for circulatory-respiratory diseases analyzed together. When circulatory diseases were analyzed alone, air pollution was not statistically significant. For this reason, only respiratory and respiratory-circulatory diseases together were analyzed in the respecified models using all the data.

In the S.I.P.S. regression of Model (8.2), air pollution was not statistically significant at the five percent level in affecting the consumption of outpatient medical services used to treat respiratory illnesses. The highest R^2 for the S.I.P.S. testing of (8.2) occurred when the air pollution and meteorological variables were lagged one day. Hence, given a severe budget constraint on the funds available for computer use, analysis of this model was not extended to other disease categories.

It was expected there might be a problem with autocorrelation in this analysis. Autocorrelation occurs when the observations on the dependent variable are not independent of each other. A plot of the residuals (the difference between the predicted observations obtained from the regression equation and the actual observations) indicated that such a problem did not exist for either of the disease categories tested.

The results of the S.I.P.S. regressions may be seen in Appendix Tables 1, 2, and 3. Attention is now turned to discussing the analysis associated with the respecified statistical models using all of the data.

The Regression Results: Statistical Model 8.1

The respecified statistical models were tested for each of several disease categories specified within the dependent variable. To be more specific, Model 8.1 was tested with respiratory diseases alone, and with circulatory-respiratory diseases combined. The regression results for each disease category can be seen in Tables 10 and 11, pages 188 and 193, respectively. Sub-Models 1 through 4 indicate the lags for the air pollution and meteorological variables. Sub-Model 1 is not lagged, Sub-Model 2 is lagged by one day, Sub-Model 3 is lagged by two days, and Sub-Model 4 is lagged by three days.

Before proceeding directly to the regression results, it would perhaps be best to summarize the characteristics assumed inherent in the dependent variable of (8.1), Y_{ijk}. All the outpatient medical services used to treat a single disease incident per contact with the medical system were expressed in California Relative Value Units. The units were transformed to dollar values such that the dollar values would approximate the <u>quantity</u> of medical services consumed per contact.

This section will discuss the regression results of Model 8.1 by independent variable. The regression results for the respiratory disease model will be presented first, then the results of the circulatory-respiratory diseases combined will be discussed.

Age of the Patient: (X_{1ijk})

Respiratory Diseases. The estimated coefficients of this variable were not significantly different from zero (see the results in Table 10). The average age of the patient consuming outpatient medical services for respiratory diseases was 46 years.

Circulatory-Respiratory Diseases. The estimated coefficient for age was significantly different from zero at the 20 percent level for Sub-Model 1 in Table 11.[72/] The coefficients in the other models were not statistically significant.

The sign of the coefficient in Model 1 was negative. The hypothesized sign for this variable, in Chapter V, was positive; i.e., the older the patient becomes, the more medical services he consumes. The negative coefficient indicates that he consumes less medical service.

Two explanations can be given for the negative sign on the coefficient. One, since the variable is barely significant at 20

[72/] Because of the nature of the study and the type of data being used, it was decided that any coefficient not significantly different from zero at, minimally, the 20 percent level would be considered not statistically significant for purposes of this research.

percent,[73/] the results should be discounted; i.e., there is a one in five chance of committing a type I error for the statistical test.

The average age for patients consuming outpatient medical services for circulatory-respiratory illnesses is 54 years, for respiratory diseases it was 46. This indicates that an older group of patients are obtaining treatment for circulatory-respiratory problems. Hence, the second reason the coefficient on the age variable may be negative is as follows. Most circulatory-respiratory diseases in older people may be severe enough to require hospitalization, such that fewer medical services are consumed on an outpatient basis. Whereas younger patients, with a probable better state of health, would most likely not be hospitalized. Treatment of the diseases for the younger patients would be at a clinic, and, proportionately, they would consume more outpatient medical services.[74/]

Sex of the Patient: (X_{2ijk})

Respiratory Diseases. The estimated coefficients of this

[73/] The critical value of t at (1,∞, 0.95) = 1.282 (for a two-tailed t test) (17, p. 305). The calculated t for Sub-Model 1, Table 11: 1.289.

[74/] This is not to argue that older patients may not visit a clinic more than younger patients. It is argued that, proportionately, they consume fewer medical services per visit.

variable were significantly different from zero at the five percent level in Model 8.1, for all lags. The signs of the coefficients were negative, which indicated that women consumer fewer medical services than men. This is not unexpected when one considers that the incidence of many respiratory diseases are lower for women than men. Some diseases are less prevalent in one sex group. If this happens, then persons within the group having the disease might be less acutely ill than individuals outside the group who also might contact the disease. Over 68 percent of the patients consuming outpatient medical services were women.

Circulatory-Respiratory Diseases. The regression coefficients were not statistically significant. This would tend to indicate that, at least with circulatory-respiratory diseases combined, sex does not have a discriminating role in the consumption of outpatient medical services. Approximately 64 percent of the patients consuming outpatient medical services were women.

Marital Status: (X_{3ijk})

Respiratory Diseases. The coefficients of the marital status variable were not statistically significant. This would tend to indicate that marital status has no effect on the consumption of medical services. Over 84 percent of the patients receiving outpatient medical services were married.

Circulatory-Respiratory Diseases. As with respiratory diseases, the marital status coefficients were not statistically different from zero. While over 81 percent of the patients receiving treatment for circulatory-respiratory diseases were married, being married seems to have no influence on their consumption of medical services.

Number of People in the Patient's Household: (X_{4ijk})

Respiratory Diseases. The estimated coefficients of this variable were not statistically significant. Although the average number of people within the patient's household was 3.39, this appeared not to influence the consumption of outpatient medical services. This does not deny the fact there may be more visits to the clinic because of the crowding. However, it would appear that crowding has no affect on the severity of the disease state.

Circulatory-Respiratory Diseases. Since crowding had no effect on the consumption of medical services with respiratory diseases, it was not surprising to find it statistically insignificant with circulatory-respiratory diseases. The average number of people in the patient's household, who are suffering from circulatory-respiratory diseases, was 2.89. Again this indicates that an older set of people were being considered in the circulatory-respiratory disease category.

Household Income: (X_{5ijk})

Respiratory Diseases. It was argued in Chapter II that the consumption of medical services, in the treatment of cardiovascular-respiratory problems, were not affected by the price of the service or the patient's income, regardless of the sources from which the care was received. The argument was supported to some extent by empirical evidence. The main reasons why the demand for medical care was assumed to be price and income inelastic were as follows: the nature of the diseases being examined; the minimum amount of medical services required to maintain the patient's state of health who is suffering from the diseases; and, with respect to price, the inability of the consumer to effectively evaluate alternative medical procedures at different prices.

From strictly a medical standpoint, price and income do not influence the Kaiser patient's ability to consume medical services; i.e., all Kaiser members, *ceteris paribus*, have essentially the same opportunity to consume identical medical services. Given these facts, the assumption of the above paragraph allowed the following assertion to be made. The use of a prepaid system, such as Kaiser's, can be used to estimate the medical costs of certain diseases associated with air pollution, and the estimates derived will be representative of other medical systems.

Household income was included in this study in an attempt to describe the patient's ability to make expenditures, in lieu of consuming medical services, which would preserve or improve his state of health. Examples of such expenditures would be those made for sound housing and nutritional food. The coefficients of the household income variables were not statistically significant in any of the Sub-Models of (8. 1).

It would appear, at least for outpatient medical services, that household income does not play a significant role in this respect. One reason for this, perhaps, is the availability of income supplements within the Portland area to low income families. These would include welfare, availability of surplus foods, and/or food stamps. The income supplements and the opportunity for all Kaiser members to consume the same kinds of medical services, would tend to minimize any differences in the consumption of medical services as between families having different income levels. The mean household income of patients consuming medical services for respiratory illnesses was approximately $12,085.

Circulatory-Respiratory Diseases. The coefficients were not significantly different from zero. The reasons for the nonsignificance are probably the same as those given for the respiratory diseases above.

The household income for patients consuming outpatient medical services for circulatory-respiratory illnesses was approximately $10,766. Again this reflects the older age group that is being examined in this disease category. Based on age, family size, and income, it would appear, as compared to the respiratory disease category, that many of the patients in the circulatory-respiratory category are retired.

Race: (X_{6ijk}, X_{7ijk})

Respiratory Diseases. The coefficients of the two race varibles were not statistically significant. This would imply that race is not important in the consumption of outpatient medical services as between Negroes (X_{6ijk}), other non-whites (X_{7ijk}), and whites. Of the contacts with the Kaiser system, which resulted in outpatient medical services being consumed, approximately 5.2 percent of the contacts were by Negroes, 1.6 percent by other non-whites, and 93.2 percent by whites.

Circulatory-Respiratory Diseases. Neither of the race coefficients were statistically significant. As with respiratory diseases it would appear race is not important in affecting the consumption of outpatient medical services. Of the contacts with the Kaiser system resulting in the consumption of medical services, approximately 3.2 percent of the contacts were by Negroes, 1.7 percent by other non-whites, and the other 95.1 percent by whites.

Physical Fitness: (X_{8ijk}, X_{9ijk})

Respiratory Diseases. Physical fitness was expressed as two qualitative variables. X_{8ijk} represented a great deal of time and energy expended by the patient in being physically fit; X_{9ijk} represented some time and energy expended by the patient in being physically fit. Statistically these were tested against the third category, where very little time and energy was expended being physically fit. It was expected that the more physically fit patient would consume fewer medical services.

With respiratory diseases, the coefficients were not significant at the 20 percent level. It would appear that physical fitness does not influence the consumption of outpatient medical services for respiratory diseases.

Of the contacts with the Kaiser system, which resulted in the consumption of medical services, 17.5 percent indicated they spent a great deal of time and energy being physically fit; 56.2 percent indicated they spent some time and energy being physically fit; 26.2 percent spent very little time and energy being physically fit.

Circulatory-Respiratory Diseases. The coefficient of X_{8ijk} was significantly different from zero at the 20 percent level in all sub-models. X_{9ijk} was statistically significant at the ten percent level in all sub-models. The signs on both of the coefficients were negative as expected. There is a suspected association between lack of physical exercise and coronary heart disease (25, p. 15), and lack of exercise and circulatory problems in general. Hence, with circulatory-respiratory diseases, it was not a surprise that the physical fitness variables were statistically significant, but not statistically significant when the respiratory diseases were analyzed alone.

Of the patients who contacted the Kaiser system and consumed outpatient medical services for circulatory-respiratory diseases, 18.6 percent, 50.3 percent, and 31.1 percent indicated they spent a great deal, some, and very little time and energy, respectively, being physically fit.

Drinking: (X_{10ijk})

Respiratory Diseases. The estimated coefficients for X_{10ijk} were significantly different from zero at the five percent level in all sub-models. The variable was a measure of how many times a year a respondent drank six drinks, eight cans of beer, or a bottle of table wine. The estimated coefficients of the variable had the expected positive signs. This would tend to support the hypothesis that heavy drinking can affect a state of health and influence the consumption of outpatient medical services.

The average number of times patients consumed alcohol excessively was 12.4 times a year. The mean reflects many zero values assigned to respondents who indicated they had not drank excessively during the 12 months prior to the household interview date. The standard deviation of the mean was 38.0.

Circulatory-Respiratory Diseases. For all the sub-models, the coefficients were statistically significant at the five percent level. All the coefficients had positive signs as expected. The average number of times per year patients indicated they drank excessively was 10.9. The standard deviation for the mean was 40.3.

Cigarette Smoking Index: (X_{11ijk})

Respiratory Diseases. While the coefficients had the right positive signs, none were significantly different from zero at the 20 percent level. The coefficient in Sub-Model 4 came the closest to being significantly different from zero at the 20 percent level. The index took a value of zero if the patient indicated he had never regularly smoked cigarettes in his life. Of the 2,439 respondents to the household interview, 991, or 40.6 percent, indicated they had not smoked regularly in the past. The mean smoking index for patients consuming medical services was 152.1 with a standard deviation of 247.8.

If any disease category would have been most affected by cigarette smoking, respiratory diseases should have been. There are a number of potential problems with the cigarette smoking index. Perhaps not enough weight in the index has been attached to the cigarette smoking habits of former smokers. The variable could be respecified to include only the number of cigarettes smoked per day, or the number of years the patient has smoked.

It is also quite possible that patients suffering from diseases most affected by cigarette smoking (chronic bronchitus and emphysema) would receive proportionately more of their medical services in the hospital. This would particularly be true in more severe cases of the disease. It could be the case, too, that once a patient has a disease affected by smoking, he may visit the doctor more often but the amount of medical services consumed per visit would remain constant. There is also the possibility that cigarette smoking (past or present) has no affect on the consumption of outpatient medical services for respiratory diseases.

Circulatory-Respiratory Diseases. The coefficients were not statistically significant in any of the models. If the coefficients were not statistically significant with respiratory diseases, the fact they were not significant with circulatory-respiratory diseases combined was not surprising. Again the lack of statistical significance may be due to the specification of the variable, the nature of or the fact that there is no association between cigarette smoking (past and present) and consumption of outpatient medical services. The mean smoking index for patients consuming medical services was 153.9 with a standard deviation of 247.2.

Occupational Exposure Index: (X_{12ijk})

Respiratory Diseases. This variable was expressed as an index. A value of zero was assigned to patients who did not work. Values greater than zero were assigned to each working patient, based upon the number of hours worked per day (plus transit time) times a weighting factor representing job related exposure to dusts.

The coefficients were not significantly different from zero at the 20 percent level. This is not surprising given the fact that 23 percent of the subscribers in the five percent sample were employed in clerical-sales (22, p. 301), resulting in a low occupational exposure index. Also approximately 39 percent of the respondents to the questionnaire did not work at all. The mean occupational exposure index of those seeking medical services was 3.4 with a standard deviation of 5.4.

Circulatory-Respiratory Diseases. The coefficients were statistically significant at the ten percent level. The signs of the coefficients were positive as expected. Why the coefficients would be statistically significant for circulatory-respiratory diseases, and not for respiratory diseases alone, is a mystery. It is possible the index may be picking up some variations in other variables not included in the model. It is not known what these other variables could be. The mean occupational exposure index for those consuming medical services was 2.6 with a standard deviation of 4.9.

Air Pollution: (X_{13ijk})

Respiratory Diseases. The coefficients of the air pollution variables were significantly different from zero for each of the sub-models in Table 10 as follows: Sub-Model 1 - significantly different from zero at the five percent level; Sub-Model 2 - significantly different from zero at the two percent level; Sub-Model 3 - statistically significant at the ten percent level; Sub-Model 4 not statistically significant at the 20 percent level. All coefficients had positive signs as expected. The null hypothesis that deteriorations in air quality causes no increase in the consumption of medical services per outpatient contact with the medical system is rejected.

The highest level of statistical significance for the air pollution coefficient occurred in Sub-Model 2, where the variable was lagged by one day. This model also had the highest R^2 of the four models.

This would tend to confirm the belief expressed earlier that there is a delay between exposure to air pollution and contact with a Kaiser clinic.

Using the coefficients, from each of the sub-models in which they were statistically significant, a 20 microgram increase in air pollution, from 60 to 80 micrograms per cubic meter, would increase outpatient medical costs per contact as follows: Sub-Model 1, an increase of 3.3 cents; Sub-Model 2, an increase of 3.5 cents; Sub-Model 3, an increase of 2.4 cents.

For the variable lagged one day, the average exposure to air pollution, of patients seeking outpatient medical services, was 61.8 micrograms per cubic meter. The standard deviation was 35.2. The means for each of the other lags may be seen in Table 10.

Circulatory-Respiratory Diseases. The air pollution coefficients were not statistically significant. The null hypothesis that deterioration in air quality causes no increase in the consumption of medical services per outpatient contact with the medical system is not rejected.

One reason, perhaps, why air pollution is not statistically significant is that many circulatory problems may be severe enough to require hospitalization. Any outpatient medical services consumed for treatment of circulatory problems would be routine, thereby limiting the variation in the dependent variable. The average unlagged exposure to air pollution for patients obtaining medical care was 62.4 micrograms. The standard deviation about the mean was 34.3.

Meteorological Conditions: (X_{14ijk})

The T.H.I. was substituted for degree days as a measure of the meteorological conditions in the study area. Originally, the main purpose for developing the T.H.I. was to have a simple method of determining the effect of summer conditions in the United States on human comfort (65, p. 41). There are several ways to calculate T.H.I. The following formula (65, p. 41) was used in this study.

$$T.H.I. = 0.4\,(T_d + T_w) + 15 \qquad (8.3)$$

where

T_d = dry-bulb (air) temperature in ^{o}F.

T_w [75/] = wet bulb temperature in ^{o}F.

75/ Atmospheric humidity can be measured by use of a mercury thermometer with a moistened wick surrounding the mercury reservoir. With adequate ventilation of the wick, its temperature is lowered by evaporation to a constant value, the wet-bulb temperature. This, together with air temperature and suitable tables, can be used to determine dew point or relative humidity (38, p. 7).

The only meteorological station in Portland which measured wet-bulb temperature was at the Portland International Airport. Hence, both wet-bulb and dry-bulb temperatures used in calculating the T.H.I. were measured at the Portland Airport, and inferred as being representative of the entire study area.

The T. H. I. was substituted for degree-days in this study because temperature and humidity have been associated with both cardiovascular and respiratory problems. T. H. I. allowed both conditions to be expressed in one variable. Empirical studies in the United States have shown that ten percent of the general population becomes uncomfortable when T. H. I. reaches a value of 70. Fifty percent of the population becomes uncomfortable when the T. H. I. reaches 75 (65, p. 42).

As previously mentioned, higher values of this index mean increased discomfort to the patient. For patients susceptible to circulatory and respiratory problems, higher T. H. I. values could aggrevate existing health conditions and result in increased consumption of medical services. Hence, it is expected the coefficient of this variable will be positively related to the dependent variable. Attention is now turned to the statistical results obtained from this variable.

Respiratory Diseases. The estimated coefficients for the T. H. I. variables were statistically significant at the one percent level for all the models in Table 10. All the coefficients had the expected positive signs. This would tend to support the supposition that the combined factors of temperature and humidity influence the consumption of medical services.

For all the sub-models, the mean T. H. I. value of those

Table 10. Regression results of statistical model 8. 1: respiratory disease.

	Sub-Model 1 - Unlagged			Sub-Model 2 - Lagged One Day		
Variables	Coefficients[a]	t Statistic	Mean[b] (Untransformed)	Coefficients[a]	t Statistic	Mean[b] (Untransformed)
Sample Size	1700			1700		
Constant	4.0995E-01			1.7224E-01		
Age	9.2984E-04	0.5271	46.0506	9.4703E-04	0.5373	c
	(1.7641E-03)		(16.0231)	(1.7625E-03)		
Sex	-9.2609E-02	2.1011	0.6818	-9.1826E-02	2.0851	c
	(4.4077E-02)		(0.5537)	(4.4040E-02)		
Marital Status	4.9017E-02	0.6946	0.8424	5.0257E-02	0.7127	c
	(7.0571E-02)		(0.3645)	(7.0520E-02)		
Number of People in Household	-2.2342E-04	0.0135	3.3906	-5.5410E-04	0.0335	c
	(1.6566E-02)		(1.7783)	(1.6546E-02)		
Household Income	1.3839E-06	0.4229	12085.4412	1.3071E-06	0.3999	c
	(3.2721E-06)		(7525.6879)	(3.2683E-06)		
Race-Negro	-1.2499E-01	1.1793	0.0524	-1.2354E-01	1.1669	c
	(1.0598E-01)		(0.2228)	(1.0587E-01)		
Race-Other Non-white	1.1757E-01	0.6240	0.0159	1.1621E-01	0.6177	c
	(1.8843E-01)		(0.1251)	(1.8814E-01)		
Physically Fit	-3.6543E-02	0.5091	0.1753	-3.6802E-02	0.5132	c
	(7.1782E-02)		(0.3803)	(7.1711E-02)		
Somewhat Physically Fit	-6.6018E-02	1.1855	0.5624	-6.7155E-02	1.2070	c
	(5.5690E-02)		(0.4962)	(5.5637E-02)		

Table 10. Continued

Variables	Sub-Model 1 - Unlagged Coefficients[a]	t Statistic	Mean[b] (Untransformed)	Sub-Model 2 - Lagged One Day Coefficients[a]	t Statistic	Mean[b] (Untransformed)
Drinking	1.4501E-03 (6.4242E-04	2.2573	12.4064 (37.9688)	1.4583E-03 (6.4184E-04)	2.2721	c
Smoking Index	1.1919E-04 (9.8823E-05)	1.2061	152.0516 (247.8353)	1.1652E-04 (9.8761E-05)	1.1798	c
Occupational Exposure Index	1.3806E-06 (4.5747E-03)	0.3018	3.4162 (5.3893)	1.5975E-03 (4.5681E-03)	0.3497	c
Air Pollution	8.1786E-02 (3.6601E-02)	2.2346	61.2532 (34.8633)	8.4906E-02 (3.5969E-02)	2.3605	61.7633 (35.2088)
Temperature Humidity Index	3.9050E-01 (1.5010E-01)	2.6017	53.0533 (8.0905)	4.4765E-01 (1.4798E-01)	3.0250	52.9020 (8.1066)
Outpatient Medical Services			14.7539 (12.0912)			
R^2	2.1991E-02			2.3885E-02		

Table 10. Continued

Variables	Sub-Model 3 - Lagged Two Days Coefficients[a]	t Statistic	Mean[b] (Untransformed)	Sub-Model 4 - Lagged Three Days Coefficients[a]	t Statistic	Mean[b] (Untransformed)
Sample Size	1700			1700		
Constant	3.2088E-01			2.8884E-01		
Age	9.8659E-04 (1.7643E-03)	0.5592	c	1.0105E-03 (1.7641E-03)	0.5728	c
Sex	-9.3276E-02 (4.4085E-02)	2.1158	c	-9.6809E-02 (4.4086E-02)	2.1959	c
Marital Status	4.6437E-02 (7.0569E-02)	0.6580	c	4.3497E-02 (7.0582E-02)	0.6163	c
Number of People in Household	-4.1014E-04 (1.6564E-02)	0.0248	c	2.3213E-04 (1.6565E-02)	0.0140	c
Household Income	1.2893E-06 (3.2721E-06)	0.3940	c	1.1461E-06 (3.2724E-06)	0.3502	c
Race-Negro	-1.2677E-01 (1.0600E-01)	1.1959	c	-1.2602E-01 (1.0601E-01)	1.1887	c
Race-Other Non-white	1.2076E-01 (1.8824E-01)	0.6415	c	1.2032E-01 (1.8821E-01)	0.6393	c
Physically Fit	-3.2048E-02 (7.1742E-02)	0.4467	c	-3.1869E-02 (7.1726E-02)	0.4443	c
Somewhat Physically Fit	-6.5781E-02 (5.5702E-02)	1.1809	c	-6.3444E-02 (5.5724E-02)	1.1385	c

Table 10. Continued

Variables	Sub-Model 3 - Lagged Two Days			Sub-Model 4 - Lagged Three Days		
	Coefficients[a]	t Statistic	Mean[b] (Untransformed)	Coefficients[a]	t Statistic	Mean[b] (Untransformed)
Drinking	1.4623E-03 (6.4238E-04)	2.2763	c	1.4348E-03 (6.4241E-04)	2.2334	c
Smoking Index	1.2433E-04 (9.8795E-05)	1.2585	c	1.2607E-04 (9.8778E-05)	1.2763	c
Occupational Exposure Index	1.7257E-03 (4.5709E-03)	0.3776	c	1.9866E-03 (4.5709E-03)	0.4346	c
Air Pollution	6.4037E-02 (3.6881E-02	1.7363	62.1665 (35.1278)	4.6301E-02 (3.6696E-02)	1.2617	62.1979 (35.2225)
Temperature Humidity Index	4.3036E-01 (1.4653E-01)	2.9370	52.9290 (8.1986)	4.5616E-01 (1.4446E-01)	3.1577	53.0916 (8.3442)
Outpatient Medical Services						
R^2	2.1868E-02			2.1725E-02		

[a]The standard error of the regression coefficient is in parentheses. E-01, E 01, E 00, etc., indicate that the decimal place of the number is to be shifted to the left by one place; shifted to the right by one place; or not shifted at all, respectively.

[b]The standard deviation about the mean is in parentheses.

[c]The mean for this variable is the same as its mean in Sub-Model 1. Only the means of the lagged variables changed per model.

consuming outpatient medical services for respiratory diseases was approximately 53.0, with the standard deviation being approximately 8.1. This does not compare to the empirical evidence citing a T.H.I. of 70 as causing ten percent of the population to become uncomfortable. However, since this study is dealing with respiratory diseases, one would expect the patients to be considerably more susceptible to temperature and humidity than would members of the general population. This would imply that the T.H.I. values for these patients could be lower than those cited in the empirical studies, and still cause discomfort to the patient.

Circulatory-Respiratory Diseases. The coefficients of the T.H.I. variable were significantly different from zero in each of the models in Table 11 as follows: Sub-Model 1 - significantly different from zero at the two percent level; Sub-Model 2 - significantly different from zero at the one percent level; Sub-Model 3 - statistically significant at the two percent level; Sub-Model 4 statistically significant at the one percent level. The signs of the coefficients were positive as expected.

For all the sub-models, the mean T.H.I. value for patient's suffering from circulatory-respiratory diseases was approximately 54.0 with a standard deviation of approximately 8.2. The same argument applies here as was used with the respiratory diseases.

Table 11. Regression results of statistical model 8.1: circulatory-respiratory diseases.

Variables	Sub-Model 1 - Unlagged Coefficients[a]	t Statistic	Mean[b] (Untransformed)	Sub-Model 2 - Lagged One Day Coefficients[a]	t Statistic	Mean[b] (Untransformed)
Sample Size	3363			3363		
Constant	1.3190E 00			1.2441 E 00		
Age	-1.6382E-03 (1.2708E-03)	1.2891	54.3717 (16.2918)	-1.6180E-03 (1.2698E-03)	1.2742	c
Sex	2.4635E-04 (2.9157E-02)	0.0084	0.6375 (0.5613)	1.3022E-04 (2.9151E-02)	0.0045	c
Marital Status	2.9929E-03 (4.4089E-02)	0.0679	0.8112 (0.3914)	2.8894E-03 (4.4082E-02)	0.0655	c
Number of People in Household	1.7845E-03 (1.3087E-02)	0.1364	2.8944 (1.6443)	1.8327E-03 (1.3082E-02)	0.1401	c
Household Income	-1.1933E-06 (2.2427E-06)	0.5321	10765.5367 (7781.8663)	-1.1682E-06 (2.2433E-06)	0.5207	c
Race-Negro	-9.6164E-02 (8.9548E-02	1.0739	0.0324 (0.1771)	-9.6459E-02 (8.9532E-02)	1.0774	c
Race-Other Non-white	8.0489E-02 (1.2691E-01)	0.6342	0.0167 (0.1280)	8.1491E-02 (1.2684E-01)	0.6425	c
Physically Fit	-6.2167E-02 (4.5984E-02)	1.3519	0.1861 (0.3893)	-6.2172E-02 (4.5976E-02)	1.3523	c
Somewhat Physically Fit	-6.0504E-02 (3.6337E-02)	1.6651	0.5031 (0.5001)	-6.0654E-02 (3.6328E-02)	1.6696	c

Table 11. Continued

	Sub-Model 1 - Unlagged			Sub-Model 2 - Lagged One Day		
Variables	Coefficients[a]	t Statistic	Mean[b] (Untransformed)	Coefficients[a]	t Statistic	Mean[b] (Untransformed)
Drinking	7.9656E-04 (4.0020E-04)	1.9904	10.8700 (40.2684)	7.9294E-04 (4.0011E-04)	1.9818	c
Smoking Index	6.0772E-05 (6.6437E-05)	0.9147	153.9242 (247.1794)	6.0766E-05 (6.6422E-05)	0.9148	c
Occupational Exposure Index	6.2953E-03 (3.4506E-03)	1.8244	2.6489 (4.8684)	6.3640E-03 (3.4489E-03)	1.8453	c
Air Pollution	2.1134E-02 (2.5430E-02)	0.8311	62.3909 (34.3420)	1.9713E-02 (2.5015E-02)	0.7881	62.6497 (34.6665)
Temperature Humidity Index	2.4138E-01 (1.0057E-01)	2.4001	54.3396 (8.1619)	2.6135E-01 (9.9166E-02)	2.6355	54.2546 (8.2262)
Outpatient Medical Services			13.9063 (12.6867)			
R^2	8.0063E-03			8.3310E-03		

Table 11. Continued

Variables	Model 3 - Lagged Two Days: Coefficients[a]	t-Statistic	Mean[b] (Untransformed)	Model 4 - Lagged Three Days: Coefficients[a]	t-Statistic	Mean[b] (Untransformed)
Sample Size	3363			3363		
Constant	1.4245E 00			1.3690E 00		
Age	-1.5595E-03 (1.2698E-03)	1.2281	c	-1.5932E-03 (1.2701E-03)	1.2544	c
Sex	-9.2519E-04 (2.9156E-02)	0.0317	c	-1.7524E-03 (2.9143 E-02)	0.0601	c
Marital Status	2.6729E-03 (4.4103E-02)	0.0606	c	1.4193E-03 (4.4101E-02)	0.0322	c
Number of People in Household	1.9317E-03 (1.3086E-02)	0.1476	c	2.0741E-03 (1.3083E-02)	0.1585	c
Household Income	-1.2578E-06 (2.2455E-06)	0.5601	c	-1.3074E-06 (2.2458E-06	0.5821	c
Race-Negro	-9.6901E-02 (8.9581E-02)	1.0817	c	-9.6205E-02 (8.9583E-02)	1.0739	c
Race-Other Non-white	8.3529E-02 (1.2687E-01)	0.6584	c	8.1589E-02 (1.2684E-01)	0.6432	c
Physically Fit	-6.1298E-02 (4.5995E-02)	1.3327	c	-6.0687E-02 (4.5995E-02)	1.3194	c
Somewhat Physically Fit	-6.1317E-02 (3.6338E-02)	1.6874	c	-6.0791E-02 (3.6335E-02)	1.6731	c

Table 11. Continued

Variables	Sub-Model 3 - Two Days Coefficients[a]	t Statistic	Mean[b] (Untransformed)	Sub-Model 4 - Three Days Coefficients[a]	t Statistic	Mean[b] (Untransformed)
Drinking	7.9628E-04 (4.0027E-04)	1.9894	c	7.9438E-04 (4.0019E-04)	1.9850	c
Smoking Index	6.2927E-05 (6.6441E-05)	0.9471	c	6.1176E-05 (6.6455E-05)	0.9206	c
Occupational Exposure Index	6.3350E-03 (3.4490E-03)	1.8368	c	6.3415E-03 (3.4484E-03)	1.8390	c
Air Pollution	-1.3033E-03 (2.5491E-02)	0.0511	62.8473 (34.6274)	-8.4495E-03 (2.5556E-02)	0.3306	62.7691 (34.5243)
Temperature Humidity Index	2.3656E-01 (9.8931E-02)	2.3912	54.2279 (8.2513)	2.5834E-01 (9.9073E-02)	2.6076	54.3663 (8.2780)
Outpatient Medical Services						
R^2	7.6485E-03			7.9576E-03		

[a]The standard error of the regression coefficient is in parentheses. E-01, E 01, E 00, etc., indicate that the decimal place of the number is to be shifted to the left by one place; shifted to the right by one place; or not shifted at all, respectively.

[b]The standard deviation about the mean is in parentheses.

[c]The mean for this variable is the same as its mean in Sub-Model 1. Only the means of the lagged variables changed per model.

One would expect that those suffering from circulatory-respiratory illnesses to be more susceptible to temperature and humidity, and hence consume more medical services.

Looking at the t statistics in Table 11, notice that T.H.I. is most significant when lagged one day. T.H.I. was most significant with respiratory diseases (Table 10) when lagged three days.

Summary

Respiratory Diseases. The following coefficients of the variables were significantly different from zero in Model (8.1) (at 20 percent or higher): sex (X_{2ijk}); drinking habits (X_{10ijk}); air pollution (X_{13ijk}), except in Sub-Model 4; and meteorological conditions - T.H.I. (X_{14ijk}). The overall F tests for regression on all four models were statistically significant at the one percent level.

Circulatory-Respiratory Diseases. In Model (8.1) the following coefficients were statistically significant at 20 percent or higher: age (X_{1ijk}), significant only in Sub-Model 1 and with a negative sign; physical fitness (X_{8ijk} and X_{9ijk}); drinking habits (X_{10ijk}); occupational exposure (X_{12ijk}); meteorological conditions T.H.I. (X_{14ijk}). The overall F tests for regression on all four models were statistically significant at the five percent level.

The Regression Results: Statistical Model 8. 2

The main difference between Models (8. 1) and (8. 2) was that (8. 2) expressed the dependent variable as the per capita costs of outpatient medical visits per census tract. Model (8. 2) was specified in an attempt to determine whether air pollution has an effect on the number of contacts with the medical system, where the cost per contact is more or less constant. If the hypothesis that air pollution has no affect on the consumption of medical services per contact was accepted, then it was felt the effects of air pollution on the number of contacts with the system could be tested indirectly with (8. 2). The main fault with (8. 2) was that only socioeconomic-demographic characteristics of patients contacting the Kaiser system were accounted for. The socioeconomic-demographic characteristics of those who were at risk (at risk to contract a disease and use the Kaiser facilities) in each census tract were not taken into account within the independent variables. Only partial account was taken of those at risk in the dependent variable, and only if there was an outpatient medical contact from that census tract.

As a result of the S.I.P.S. regressions, only respiratory diseases were specified within the dependent variable. Given a severe budget constraint on the funds available for computer use, Model (8. 2) was tested using only a one day lag. The regression results of these models may be seen in Table 12, which also contains the means and standard deviations of the variables used in the model.

The following variables were statistically different from zero at the 20 percent level, or higher, in statistical Model (8. 2): sex ($\bar{X}_{2icj}$); household income ($\bar{X}_{5icj}$), which had a positive relationship with the dependent variable instead of the negative relationship originally hypothesized; drinking habits ($\bar{X}_{7icj}$); smoking ($\bar{X}_{8icj}$), and T. H. I. ($\bar{X}_{10icj}$). The overall F test for regression was barely statistically significant at the five percent level.

The results from Model (8. 2) should be interpreted with caution, primarily because it appears the dependent variable lacks variability. One reason for this was that there were too few multiple observations per census tract per day resulting in the consumption of outpatient medical services. This is illustrated in Table 10, which shows 1, 700 patients with respiratory illnesses who contacted the Kaiser system and consumed outpatient medical services. Table 12 shows that even after aggregation over the census tracts for the two year study period, there were still 1, 569 observations that entered regression. This means that during the two year study period only 131 observations came from the same census tract on the same day. The other 1, 569 observations were single observations per census tract. Hence, in most instances the independent variables were individual patient characteristics and not aggregate

Table 12. Regression results of statistical model 8.2: respiratory diseases - one day lag.

Variables	Coefficients[a]	t - Statistic	Mean[b] -Untransformed
Sample Size	1569		
Constant	-2.4077E-00		
Age	1.0900E-03 (2.2164E-03)	0.4918	46.0153 (16.0014)
Sex	9.6563E-04 (5.9370E-04)	1.6265	65.6474 (51.8897)
Marital Status	-4.7287E-04 (8.9162E-04)	0.5303	84.4806 (36.1985)
Number of People in Household	3.5658E-03 (2.0714E-02)	0.1721	3.3996 (1.7804)
Household Income	6.9822E-06 (4.1046E-06)	1.7011	12112.6119 (7540.2050)
Race-Non-white	3.1119E-04 (1.1727E-03)	0.2654	7.0746 (25.6481)
Drinking	1.0742E-03 (7.8481E-04)	1.3687	12.8749 (39.1389)
Smoking Index	1.7571E-04 (1.2237E-04)	1.4359	154.0181 (250.5819)
Occupational Exposure Index	4.0458E-03 (5. 7227E-03)	0.7070	3.4413 (5.4109)
Temperature Humidity Index	4.6304E-01 (1.8806E-01)	2.4621	53.0149 (8.0738)
Air Pollution	4.5315E-02 (4.5460E-02)	0.9968	61.7121 (34.9227)
Per Capita Costs			1.6083 (2.4151)
R^2	1.1830E-02		

[a]The standard error of the regression coefficient is in parentheses. E-01, E 01, E 00, etc., indicate that the decimal place of the number is to be shifted to the left by one place; shifted to the right by one place; or not shifted at all, respectively.

[b]The standard deviation about the mean is in parentheses.

data as originally specified. Having more observations per census tract might have afforded more variability within census tracts. Lack of sufficient number of observations per census tract prohibited testing whether air pollution has an affect on the number of contacts with the medical system.

As mentioned in Chapter II, even if Model (2. 5) had been correctly specified, such that the null hypothesis Ho_2 (deterioration in air quality does not precipitate an increase in the number of contacts with the medical system per disease category) could have been tested, there were not enough multiple observations per census tract to get a meaningful test. That is, there would have been the same problem there was with the dependent variable in (8. 2) - not enough variation.

Another possibility for lack of variation in the dependent variable, as contributed by the outpatient medical costs per census tract, would be the type of medical services being examined by this study. This will be taken up again in the next chapter. Suffice it to say that many outpatient contacts with the medical system may be of a routine nature, which would not provide much of an opportunity for variation in the consumption of medical services per contact. If this is true, it may be that any variation in the dependent variable of (8. 2) is contributed by the denominator, which is the number of people living in the census tract from which a medical contact was made.

IX. SUMMARY, CONCLUSIONS, RECOMMENDATIONS

Summary

The main supposition of this research was that air pollution can aggravate a state of health resulting in increased consumption of outpatient medical services, and the number of contacts with the medical system, for certain respiratory, cardiovascular, and other diseases aggravated by air pollution. The main objective of this study was to quantify such effects, if any, and express them in monetary terms.

The study examines the effect of day to day changes in air quality on the consumption of outpatient medical services. It was felt that, in order to isolate the effect of air pollution on health, account must be taken of other variables within the explanatory models which could influence the consumption of medical services. These included socioeconomic-demographic and meteorological variables.

The study period was for 1969 and 1970. The study area was the Portland, Oregon-Vancouver, Washington area. Kaiser Health Services Research Center, associated with Kaiser Foundation Health Plan in Portland, Oregon (a group prepaid medical plan with 150,000 members), provided the outpatient medical and socioeconomic-demographic data. The outpatient medical records were obtained from a five percent, ongoing, random sample of the total Kaiser membership. The socioeconomic-demographic data came from a questionnaire administered by Kaiser to the sample members during 1970 and 1971.

The air quality data used in this study came from the following Oregon and Washington air pollution control authorities: Columbia-Willamette Air Pollution Authority, Department of Environmental Qualtiy, and the Southwest Air Pollution Control Authority. The measure of air quality used in this study was suspended particulate matter. The meteorological data were provided by the National Weather Service.

It is argued that in the consumption of medical services for cardiovascular-respiratory problems, the price of medical services and the patient's income may be discounted as variables affecting the demand for medical care, regardless of the medical care source. The argument is supported by empirical evidence. The basis of support for the argument, for purposes of this study, are as follows: the nature of the diseases being examined (without medical treatment certain cardiovascular-respiratory diseases will terminate in death); a minimum amount of medical services required to maintain the patient's health state who is suffering from the diseases; with respect to price, the inability of the patient to effectively evaluate alternative medical procedures at different prices. Hence, it is concluded, the use of a prepaid system such as Kaiser's can be used to estimate the medical costs of certain diseases associated with air pollution, and the estimates derived will be respresentative of other medical systems.

All the medical services used to treat a single disease incident per clinic visit, or other contact with the medical system, were expressed in California Relative Value (C. R. V.) units. The units reflected the best approximation of the amount of physician output actually provided in rendering medical services to the patient. The units can be converted to dollar value and still have an accurate approximation of the quantity of medical services consumed.

A list was formulated of over 500 diseases and symptoms thought to be associated with air pollution. The list was divided into eight main disease categories. The respiratory and circulatory disease categories were used to pretest two statistical models formulated for this study. A ten percent, systematic, sample was drawn from the medical records to test the models. The purpose of the pretesting was to gain insight into the transformations required on the different variables, and to minimize computer expense. As a result of the pretests the regression models were respecified, and were labeled statistical Models (8. 1) and (8. 2).

The consumption of outpatient medical services for respiratory and circulatory-respiratory diseases were specified separately within the dependent variable of Model (8. 1). The air pollution and meteorological variables were lagged up to three days to reflect the fact there may be a time sequence in casuality between exposure to air pollution and meteorological phenomena, and the onset of an illness. A separate regression model evolved for each lagged day in (8. 1) and (8. 2). These were labeled sub-models, where Sub-Model 1 represented the unlagged day, etc. ,up through Sub-Model 4 which represented a three day lag. The regression results for each sub-model may be seen in Tables 10 and 11 in Chapter VIII. The statistical results are summarized as follows.

Respiratory Diseases. Except where noted, in all four sub-models the following coefficients were significantly different from zero at the 20 percent level, or higher: sex (X_{2ijk}), negatively related to the dependent variable; drinking habits (X_{10ijk}), positively related to the dependent variable; air pollution (X_{13ijk}), the coefficients were positively related to the dependent variable in all models, but was not statistically significant in Sub-Model 4; and meteorological conditions (X_{14ijk}), positively related to the dependent variable.

The coefficients of multiple determination (R^2) for Sub-Models 1-4 were 0. 022, 0. 024, 0. 022, and 0. 022, respectively. The overall F tests for regression on the four models were statistically significant at the one percent level.

Circulatory-Respiratory Diseases. The following coefficients were statistically significant at the 20 percent level or higher, in all models, except where noted: age (X_{1ijk}), significant only in Sub-Model 1 and with a negative instead of the expected positive sign; physical fitness (X_{8ijk} and X_{9ijk}), negatively related to the dependent variable; drinking habits (X_{10ijk}), positively related to the dependent variable; occupational exposure index (X_{12ijk}), positively related to the dependent variable; meteorological conditions (X_{14ijk}), positively related to the dependent variable.

The R^2's for each of the Sub-Models 1-4 were 0.0080, 0.0083, 0.0076, and 0.0080, respectively. The overall F tests for regression on the four models were statistically significant at the five percent level.

Air pollution may increase the number of contacts with the medical system without necessarily increasing the amount of medical services consumed per office visit. Model (8.2) was specified to take account of this fact. As a result of the preliminary tests, Model (8.2) was tested only with respiratory diseases and for a lag of one day. The regression results may be seen in Table 12. The following variables were significantly different from zero at 20 percent, or higher: sex ($\overline{X}_{2icj}$) positively related to the dependent variable; household income ($\overline{X}_{5icj}$) positively related to the dependent variable instead of the expected negative relationship; drinking habits ($\overline{X}_{7icj}$), related positively to the dependent variable; smoking index ($\overline{X}_{8icj}$), this coefficient is positively related to the dependent variable; meteorological conditions ($\overline{X}_{10icj}$), positively related to the dependent variable. The overall F test for regression was barely statistically significant at the five percent level.

The results of (8.2) should be interpreted with caution, primarily because the dependent variable appeared to be limited in the amount of variation which it can exhibit. The reasons for this were thought to be as follows: One, few multiple observations per census tract per day resulted in the consumption of outpatient medical services. Two, the possibility of a limited amount of medical services being performed per contact with the medical system. Both possibilities tend to limit the amount of variation in the dependent variable to its denominator; that is, to changes in the number of people living in the different census tracts.

Limitations of the Study and the Conclusions Derived

The above study infers that air pollution does have some effect on the consumption of outpatient medical services used to treat respiratory illnesses. The results would indicate that the procedures outlined in this study appear to hold some promise in quantifying such effects. However, it is quite evident that the results presented here are only a beginning, and much work needs to be done. This section is a discussion of some of the things that appeared to be wrong with the procedures used in the study. Some suggestions are made as to how they might be corrected.

The coefficient of multiple determination (R^2) measures the total variation about the mean of the dependent variable which is explained by regression (17, p. 26). The higher the value of the R^2, the better the fitted equation explains the variation in the data (17, p. 117). The R^2's for each of the regression models in this study are extremely low. This means that the statistical models are explaining very little of the variation in dependent variables. This could imply two things: (1) The models are inadequate and should be respecified. (2) To begin with, there is too little variation in the dependent variable, and the models are not that inadequate.

The theoretical framework and the statistical models derived from it were a simple description of a complex phenomenon. The consumption of medical services, or contacts with the medical system, are difficult to explain. The consumption of medical services is not only dependent on happenings in the present time period, but also on occurrences which may have been in the patient's past. Medical care consumption can also be influenced by hereditary factors. Specifying a model to explain current consumption of medical services by observing changes, as between individuals, or socioeconomic-demographic, meteorological, and air pollution conditions is a start, but leaves much to be desired. It is apparent that the variables within the models can be respecified in many different ways. It is also quite evident that numerous other models could be specified and tested with the data available.

Another question which must be raised is whether the statistical models have been adequately tested by the data available; i. e., are outpatient medical costs really the best way to determine the medical costs of air pollution? One can argue that there are a limited amount of medical services that can be consumed on an outpatient visit. Many such visits are routine, and, while air pollution may affect the number of visits, it does not dramatically affect the cost per visit. The results of this study would tend to confirm this. This is why the effect of air pollution on the number of contacts with the medical system is so important. One of the biggest shortcomings of the data base was that there were not enough contacts with the medical system per census tract to test whether air pollution does indeed affect it.

A hint as to whether many of the outpatient visits were routine can be obtained by examining the actual observations on the dependent variable. This was accomplished by listing the Y_{ijk}, Model (8. 1), for the ten percent sample. While there was the opportunity, as per the California Relative Value (C. R. V.) tape obtained from Kaiser Research, for a variety of dollar combinations to be listed, many of the same dollar values repeated themselves consistently. This would limit the variability of the dependent variable. The repetitiveness of many of the dollar values would indicate two things. One, many of the outpatient contacts are indeed routine. Two, the coding technicians at Kaiser Research are not coding many of the procedures which would most adequately reflect the medical services consumed per contact. There is no way of checking either contention without listing the procedure numbers per contact, and comparing what is listed for each procedure number with what is listed on the outpatient medical chart.

One cannot finish this section without discussing the adequacy of the C.R.V. in reflecting the output of medical services, particularly as it refers to outpatient medical services. There are a variety of different medical services that can be provided within a given office call. To adequately reflect the multitude of services, the C.R.V. (Kaiser Adapted), while detailed, would have to be infinitely more so than they now are. Each office procedure and medical service would have to be studied in detail, and weights attached based on physician output of medical services. This would require a detailed man-power study. Kaiser Research personnel were thinking of undertaking such a study, which might help in more detail coding of outpatient medical services.

Many issues have been raised here, which cannot be answered without detailed, further study of many months. However, this does not preclude some general conclusions being reached about this research. The regression results of Model (8. 1) - respiratory disease category - showed that many of the variables expected to have an effect on the consumption of medical services did so, and with the expected relationships to the dependent variable. The same happened with the circulatory-respiratory regression of Model (8. 1), even though air pollution was not statistically significant. The results did indicate that, while air pollution does not have a dramatic effect on increasing the consumption of outpatient medical services, it does have some effect. Given the nature of outpatient medical services, as discussed above, this need not be unexpected.

While there was limited variation in the dependent variable, which caused the results from Model (8. 2) to be regarded with suspicion, too many things came out as expected with (8. 1) to consider the results of this study to be spurious. It is felt that this study represents a good beginning in a field where nothing existed before. And that the results show promise, given the limitations previously detailed. It would appear that further study of this issue would be fruitful. It is in this vein that attention is now turned to recommendations for future research.

Recommendations for Future Research

The following recommendations are made with respect to future research.

(1) Attempts should be made to investigate the questions raised in the last section before any further research using the Kaiser data base is undertaken.

(2) Assuming satisfactory answers to these questions, it is recommended that the effect of air pollution on the number of contacts with the Kaiser system be determined. This in essence means that more data beyond that contained in the five percent sample will have to be collected, such as from the entire Kaiser membership. The first problem to overcome in doing this would be obtaining the necessary outpatient medical data. The second problem to overcome would be obtaining the addresses for the patients in order that air pollution and meteorological values can be assigned to them. All the addresses of the entire Kaiser membership are supposedly going to be run through the Geographic Base File of Admatch. If this is true, the second problem will be solved, and coordinates will be available for the entire Kaiser population.

(3) It is recommended that the models hypothesized for this study be retested by respecifying some of the independent variables

which were not statistically significant.

(4) The air pollution data for the City of Portland is only adequate from the beginning of 1969. It is recommended that additional air pollution data, and outpatient medical data from the five percent sample, be collected for the years after 1970. This data should be used to retest the models postulated for this study, or other models which might be postulated as a result of (3) above or the insights of other researchers.

(5) It is recommended that steps be taken to correct the inpatient medical tape. Kaiser Research Center is currently in the process of correcting many of the errors associated with the tape. The biggest problems for future researchers to overcome are assigning costs to the different medical procedures, and reconciling the differences in the International Classification of Diseases, Adapted (I.C.D.A.) resulting from using two different editions of the code. The inpatient tape offers an excellent opportunity to study the effects of air pollution on the consumption of inpatient medical services. There should be more variety in the types of inpatient services consumed, and hence more variability in the dependent variable. The socioeconomic-demographic data are available on the magnetic tape with each inpatient record. These data are current at the time of inpatient admission to the hospital.

(6) It is suggested that any future study on outpatient-inpatient medical costs take account of drugs. Not to do so leaves out a very important medical cost. Kaiser Research has started coding the consumption of drugs at the clinic for the five percent sample beginning January 1971. It is not known if prescription drugs are also being coded. Consumption of drugs while in the hospital are not currently being coded. However, a sampling framework could be developed in which selected inpatient medical charts are examined, and the drugs consumed recorded by I. C. D. A. classifications. The results from the sample could then be inferred to the population. The same procedure could be used with the outpatient medical records. The prescriptions are usually recorded on the outpatient medical chart.

BIBLIOGRAPHY

1. American Association for the Advancement of Science. Air Conservation Commission. Air conservation. Baltimore, Horn-Shafer, 1965. 324 p. (Publication no. 85)

2. Andersen, Ronald, and Lee Benham. Factors affecting the relationship between family income and medical care consumption. In: Empirical studies in health economics. ed. by Herbert E. Klarman. Baltimore, Johns Hopkins Press, 1970. p. 73-95.

3. Auerbach, Oscar and Steven M. Spencer. New hope for heavy smokers. Readers Digest 96:129-131. Feb. 1970.

4. Berg, Neil J. and John I. Kowalczyk. An air quality index designed to serve the needs of a regional air pollution control authority. Paper presented at the annual meeting of the Pacific Northwest International Section of the Air Pollution Control Association, Portland, Oregon, November 23-25, 1969. 11 numb. leaves.

5. California Medical Association. Committee on Fees of the Commission on Medical Services. 1964 California relative value studies. 4th ed. San Francisco, Six Ninety Sutter Publications, 1964. 79 p.

6. California Medical Association. Committee on Fees of the Commission on Medical Services. 1969 California relative value studies. 5th ed. San Francisco, Six Ninety Sutter Publications, 1969. 135 p.

7. Cassell, Eric J. Epidemiology of air pollution in urban population. Paper presented at Air Pollution and Respiratory Disease: A Progress Report, St. Vincent's Hospital and Medical Center of New York, New York City, Oct. 28, 1966. 18 numb. leaves.

8. Cassell, Eric J. The health effects of air pollution and their implications for control. Law and Contemporary Problems 33: 197-216. 1968.

9. Cassell, Eric J. et al. Air pollution, weather, and illness in a New York population. Archives of Environmental Health 18: 523-530. 1969.

10. Coast and Geodetic Survey. Plane coordinate projections tables, Oregon. Washington, D. C., U. S. Government Printing Office, 1952. 31 p. (Special publication no. 270)

11. Coast and Geodetic Survey. Plane coordinate projection tables, Washington. Washington, D. C., U. S. Government Printing Office, 1952. 28 p. (Special publication no. 271)

12. Coast and Geodetic Survey. Sines, cosines, and tangents: ten decimal places with ten-second interval. 0° -6°. (For plane coordinate computation on Lambert projections.) Washington, D. C., U. S. Government Printing Office, 1949. 36 p. (Special publication no. 246)

13. Columbia Basin Inter-Agency Committee. Meteorology Subcommittee. Bibliography of meteorological data on punched cards and magnetic tape Columbia Basin States. Portland, 1965. 108 p.

14. Columbia Basin Inter-Agency Committee. Meteorology Subcommittee. Bibliography of published climatological data Columbia Basin States. Portland, 1965. 310 p.

15. Committee of the Oregon Thoracic Society. Chronic obstructive pulmonary disease: a manual for physicians. Rev. ed. Portland, Oregon Thoracic Society, 1966. 83 p.

16. Dorland's illustrated medical dictionary. 24 ed. Philadelphia, W. B. Saunders Company, 1965. 1724 p.

17. Draper, N. R. and H. Smith. Applied regression analysis. New York, Wiley, 1966. 402 p.

18. Feldstein, Paul J. The demand for medical care. In: Report of the commission on the cost of medical care. Chicago, American Medical Association, Vol. I, 1964. p. 57-76.

19. Greenburg, Leonard et al. Intermittent air pollution episodes in New York City. 1962. Public Health Reports 78:1061-1064. 1963.

20. Greenlick, Merwyn R. and Ernest W. Saward, M. D. Impact of a reduced charge drug benefit in a prepaid group practice plan. Public Health Reports 81:938-940. 1966.

21. Greenlick, Merwin R. Principal Investigator. Supplemental progress report: medical care utilization project. June 1966 to August 1968. 19 numb. leaves. (Kaiser Foundation Research Institute. Research Grant Number: 5 R01 CH 00225-03 HSR.)

22. Greenlick, Merwin R. et al. Determinants of medical care utilization. Health Services Research Winter:296-315. 1968.

23. Griffiths, John F. Applied climatology, an introduction. London, Oxford University Press, 1966. 118 p.

24. Hagstrom, R. M., H. A. Sprague, and E. Landau. The Nashville air pollution study: VII. Mortality from cancer in relation to air pollution. Archives of Environmental Health 15:237-248. 1967.

25. Hayman, Charles, M. S., and G. L. Dunnahoo, M. D. Cardiovascular disease: a public health problem. In: Cardiovascular disease, new perspectives: continuing education in public health. San Francisco, Western Regional Office, American Public Health Association, March 1962. (Continuing Education Monograph no. 2)

26. Hodgson, Thomas A., Jr. Short-term effects of air pollution and mortality in New York City. Environmental Science and Technology 4:589-597. 1970.

27. Hopkins, C. E. et al. Intrafamily correlation and its significance in the interpretation of sample survey. American Journal of Public Health 53:1112-1120. 1963.

28. Hurtado, Arnold W. and Merwyn R. Greenlick. Approaches and techniques: a disease classification system for analysis of medical care utilization, with a note on symptom classification. Health Services Research Fall:235-250. 1971.

29. Hurtado, Arnold W. and Merwyn R. Greenlick. The clinical-behavior disease classification system. Kaiser Foundation Hospitals, Portland, Oregon, n. d.

30. Jaksch, John A. and H. H. Stoevener. Effects of air pollution on residential property values in Toledo, Oregon. Corvallis, Agricultural Experiment Station, Oregon State University, September 1970. 15 p.

31. Joint Committee of the Allergy Foundation of American and the American Thoracic Society. Asthma: a practical guide for physicians. U. S. A., National Tuberculosis and Respiratory Disease Association in cooperation with the Allergy Foundation of America, 1968. 47 p.

32. Kaiser. Unpublished: Recording and coding instructions for medical care utilization study. Kaiser Foundation Hospitals Medical Care Research Unit. Rev. ed. August 1968.

33. Kilburn, Kaye H. Cilia and mucus transport as determinants of the response of lung to air pollutants. Archives of Environmental Health 14: 77-91. 1967.

34. Kissam, Philip. Surveying for civil engineers. New York, McGraw-Hill, 1956. 716 p.

35. Klarman, Herbert E. The economics of health. New York, Columbia University Press, 1965. 200 p.

36. Kutschenreuter, Paul Herbert. A study of the effect of weather on mortality in New York City. Master's thesis. New Brunswick, Rutgers, 1960. (Cited in: Hodgson, Thomas A., Jr. Short-term effects of air pollution on mortality in New York City. Environmental Science and Technology 4:589-597. 1970. p. 590.)

37. Lave, Lester B. and Eugene P. Seskin. Air pollution and human health. Science 169:723-733. 1970.

38. Licht, Sidney (ed.). Medical climatology. New Haven, Elizabeth Licht, 1964. 753 p. (Vol. 8)

39. McCarroll, James R. et al. Health and the urban environment. V. Air pollution and illness in a normal urban population. Archives of Environmental Health 15:178-184. 1967.

40. Meade, Buford R. The practical use of the Oregon state plane coordinates system. In: Proceedings of the 1964 surveying and mapping conference. Corvallis, Oregon, Engineering Experiment Station, Oregon State University, 1965. p. 5-33. (Circular no. 33)

41. Michelson, Irving and Boris Tourin. Comparative method for studying costs of air pollution. Public Health Reports 81:505-512. 1966.

42. National Center for Health Statistics. Eighth revision international classification of diseases, adapted for use in the United States. Washington, D.C., Department of Health, Education, and Welfare, Public Health Service Publication No. 1693. 1968. 2 vols.

43. National Center for Health Statistics. International classifications of diseases, adapted: for indexing hospital records by diseases and operations. Rev. ed. Washington, D.C., Department of Health, Education, and Welfare, Public Health Service Publication No. 719. 1962. 2 vols.

44. National Tuberculosis and Respiratory Disease Association. Air pollution primer. New York, 1969. 104 p.

45. National Tuberculosis Association. Introduction to respiratory diseases. 2nd. ed. U.S.A., 1964. 122 p.

46. Olsson, Lars E. and James S. Sasser. Available meteorological data and data handling facilities at Oregon State University. Corvallis, Oregon, April 1971. 28 p. (Oregon State University. Department of Atmospheric Sciences and Air Resources Center. Technical report no. 71-1)

47. Olsson, Lars E., et al. The study of the natural ventilation of the Columbia-Willamette Valleys: II. Corvallis, Oregon, April 1971. 164 p. (Oregon State University. Departments of Atmospheric Sciences and Oceanography and Air Resources Center. Technical report no. 71-2).

48. Oregon. Department of Environmental Quality. Air Quality Control. Report on Oregon's air quality and program for years 1968 and 1969. Portland, December 1969. 91 p.

49. Oregon-Washington Air Quality Committee. Criteria for particle fallout objectives and standards. Portland, Oregon, State Sanitary Authority, April 1969. 13 p.

50. Oregon-Washington Air Quality Committee. Criteria for suspended particulate objectives and standards. Portland, Oregon, State Sanitary Authority, April 1969. 12 p.

51. Painter, John, Jr. Indigents' treatment stirs key issue at Multnomah Hospital. The Sunday Oregonian (Portland, Oregon) p. 36, col. 1-8. April 18, 1971.

52. Pennington, Ralph H. Introductory computer methods and numerical analysis. New York, Macmillan, 1965. 452 p.

53. Polk's Beaverton (Washington County, Oregon) city directory: 1970. Monterey Park, California, R. L. Polk, (1970). Various paging.

54. Polk's Portland east suburban (Multnomah County, Oregon) directory: 1970. Monterey Park, California, R. L. Polk, (1970). Various paging.

55. Polk's Portland (Multnomah County, Oregon) city directory: 1970. Monterey Park, California, R. L. Polk, (1970). Various paging.

56. Polk's Vancouver (Clark County, Washington) city directory: 1969. Monterey Park, California, R. L. Polk, (1969). Various paging.

57. Portland. In: The Encyclopedia Americana. International ed. Vol. 22. New York, 1970. p. 394-406.

58. Ridker, Ronald G. Economic costs of air pollution, studies in measurement. New York, Frederick A. Praeger, 1967. 214 p.

59. Robinson, Arthur H. and Randall D. Sale. Elements of cartography. 3rd. ed. New York, John Wiley and Sons, 1969. 415 p.

60. Sterling, T. D., S. V. Pollack, and J. J. Phair. Urban hospital morbidity and air pollution. Archives of Environmental Health 15:362-374. 1967.

61. Stevens, Carl M. Physician supply and national health care goals. Industrial Relations 10:119-144. 1971.

62. The health factory: mammoth prepaid plan trims medical costs, wins praise for care. The Wall Street Journal. San Francisco ed. p. 1, col. 6. April 26, 1971.

63. Thompson, Donovan J. et al. Health and the urban environment: VIII. Air pollution, weather, and the common cold. American Journal of Public Health 60:731-737. 1970.

64. Tromp, S. W. The geographical distribution of arterio-sclerotic heart diseases in the Netherlands (period 1936-1945). Leiden, Bioclimatological Research Center, Nov. 20, 1958. (Cited in: Tromp, S. W. Medical biometeorology: weather, climate, and the living organism. Amsterdam, Elsevier, 1963. 991 p.)

65. Tromp, S. W. Medical biometeorology: weather, climate, and the living organism. Amsterdam, Elsevier, 1963. 991 p.

66. U. S. Bureau of the Census. Census of housing: 1970 block statistics, Portland, Oregon-Washington urbanized area. Final Report HC(3)-195 Washington, D.C., U. S. Government Printing Office, 1971. Various paging.

67. U. S. Bureau of the Census. 1960. Census of population, alphabetical index of occupations and industries. Rev. ed. Washington, D.C., U. S. Government Printing Office, 1960. 649 p.

68. U. S. Bureau of the Census. 1970 Census of population: general social and economic characteristics. Final Report PC(1)-C39 Oregon. Washington, D.C., U. S. Government Printing Office, 1972. Various paging.

69. U. S. Bureau of the Census. Census use study: admatch users manual. Washington, D.C., 1970. 61 p.

70. U. S. Department of Commerce. Environmental Data Service. Climatological data: Oregon. Vol. 73, No. 1 through Vol. 76, No. 12. (All months inclusive from January 1967 through December 1970)

71. U. S. Department of Commerce. Environmental Data Service. Climatological data: Washington. Vol. 71, no. 1 through Vol. 74, no. 12. (All months inclusive from January 1967 through December 1970)

72. U. S. Department of Commerce. Environmental Data Service. Local climatological data: Portland, Oregon International Airport. Washington, D.C., U. S. Government Printing Office, 1967-1970. n.p.

73. U. S. Department of Commerce. National Oceanic and Atmospheric Administration. Environmental Data Service. Local climatological data: annual summary with comparative data, Portland, Oregon. Washington, D. C., U. S. Government Printing Office, 1970. n. p.

74. U. S. Postal Service. National zip code directory: 1970-1971. Washington, D. C., U. S. Government Printing Office, 1970. 1733 p.

75. U. S. Public Health Service. Consumer Protection and Environmental Health Service. Air quality criteria for particulate matter. Washington, D. C., National Air Pollution Control Administration, January 1969, 211 p. (National Air Pollution Control Administration Publication no. AP-49)

76. U. S. Public Health Service. Consumer Protection and Environmental Health Service. Air quality criteria for sulfur oxides. Washington, D. C. National Air Pollution Control Administration, January 1969. 178 p. (National Air Pollution Control Administration no. AP-50)

77. U. S. Public Health Service. Division of Air Pollution. Air pollution measurements of the national air sampling network: analysis of suspended particulates, 1957-1961. Washington, D. C., Robert A. Taft Sanitary Engineering Center, 1962. 217 p. (Public Health Service Publication no. 978)

78. U. S. Public Health Service. Environmental Health Service. Air quality criteria for carbon monoxide. Washington, D. C., National Air Pollution Control Administration, March 1970. Various paging. (National Air Pollution Control Administration Publication no. AP-62)

79. U. S. Public Health Service. Environmental Health Service. Air quality criteria for hydrocarbons. Washington, D. C., National Air Pollution Control Administration, March 1970. Various paging. (National Air Pollution Control Administration Publication no. AP-64)

80. U. S. Public Health Service. Environmental Health Service. Air quality criteria for photochemical oxidants. Washington, D. C., National Air Pollution Control Administration, March 1970. Various paging. (National Air Pollution Control Administration Publication no. AP-63)

81. U. S. Public Health Service. Health Services and Mental Health Administration. Chart book on smoking, tobacco, and health. Arlington National Clearinghouse for Smoking and Health, 1969. 46 p. (Public Health Service Publication no. 1937)

82. U. S. Public Health Service. Health Services and Mental Health Administration. The health consequences of smoking, a report of the surgeon general: 1971. Washington, D.C., U. S. Government Printing Office. 158 p. (Department of Health, Education, and Welfare Publication no. HSM 71-7513)

83. Wendroff, Burton. Theoretical numerical analysis. New York, Academic Press, 1966. 239 p.

84. Winkelstein, Warren, Jr. et al. The relationship of air pollution and economic status to total mortality and selected respiratory system mortality in men. Archives of Environmental Health 14:162-171. 1967.

85. Zeidberg, L. D., R. J. M. Horton, and E. Landau. The Nashville air pollution study: VI. Cardiovascular disease mortality in relation to air pollution. Archives of Environmental Health 15:225-236. 1967.

APPENDIX

Appendix Table 1. S. I. P. S. regression results of statistical model 8. 1: respiratory diseases - unlagged.

Variables	Coefficients[a]	t Statistic	Mean[b] - Untransformed
Sample Size	153		
Constant	-1.8736E 00	0.9738	
	(1.9241E 00)		
Age	-6.7890E-05	0.0114	45.5294
	(5.9527E-03)		(16.0538)
Sex	1.1035E-01	0.7659	0.7386
	(1.4408E-01)		(0.5823)
Marital Status	2.8038E-01	1.2737	0.8039
	(2.2013E-01)		(0.3983)
Number of People in Household	2.4386E-02	0.4796	3.5033
	(5.0846E-02)		(1.9639)
Household Income	-7.7191E-06	0.7615	12178.1046
	(1.0137E-05)		(7869.8121)
Race-Negro	-2.4771E-01	0.7275	0.0523
	(3.4051E-01)		(0.2233)
Race-Other Non-white	4.6733E-01	0.6983	0.0131
	(6.6923E-01)		(0.1140)
Physically Fit	-1.9884E-01	0.8097	0.1307
	(2.4558E-01)		(0.3382)
Somewhat Physically Fit	-4.8518E-02	0.2900	0.5556
	(1.6731E-01)		(0.4985)
Drinking	-4.8631E-04	0.2300	13.5157
	(2.1145E-03)		(38.2885)
Smoking Index	4.5499E-04	1.1181	118.9319
	(4.0694E-04)		(198.2665)
Occupational Exposure Index	7.7237E-03	0.5570	4.3030
	(1.3866E-02)		(5.9661)
Air Pollution	2.7269E-01	2.2427	60.2792
	(1.2159E-01)		(32.0161)
Temperature Humidity Index	7.0107E-01	1.5511	52.7171
	(4.5198E-01)		(8.5303)
Outpatient Medical Services			13.8386
			(11.2557)
R^2	0.1156		

[a]The standard error of the regression coefficient is in parentheses. E-01, E 01, E 00, etc., indicate that the decimal place of the number is to be shifted to the left by one place; shifted to the right by one place; or not shifted at all, respectively.

[b]The standard deviation about the mean is in parentheses.

Appendix Table 2. S. I. P. S. regression results of statistical model 8.1: circulatory-respiratory diseases - unlagged.

Variables	Coefficients[a]	t Statistic	Mean[b] - Untransformed
Sample Size	333		
Constant	-5.9326E-01 (1.2102E 00)	0.4902	
Age	9.6186E-04 (3.9557E-03)	0.2432	54.4925 (16.0471)
Sex	2.3956E-01 (8.3891E-02)	2.8557	0.6667 (0.5860)
Marital Status	2.2654E-01 (1.2790E-01)	1.7713	0.8018 (0.3992)
Number of People in Household	3.5947E-02 (3.8202E-02)	0.9410	2.9399 (1.7552)
Household Income	-9.1927E-06 (6.5999E-06)	1.3929	10704.9550 (7900.4679)
Race-Negro	-2.1188E-01 (2.6783E-01)	0.7911	0.0300 (0.1709)
Race-Other Non-white	-5.1321E-03 (3.3190E-01)	0.0155	0.0210 (0.1437)
Physically Fit	-3.1567E-02 (1.3458E-01)	0.2346	0.1772 (0.3824)
Somewhat Physically Fit	3.4825E-02 (1.0458E-01)	0.3330	0.4955 (0.5007)
Drinking	-6.9087E-04 (1.0267E-03)	0.6729	12.3330 (46.1372)
Smoking Index	3.5109E-04 (2.1380 E-04)	1.6421	136.7889 (224.4347)
Occupational Exposure Index	1.4590E-02 (9.6653E-03	1.5095	3.1037 (5.2676)
Air Pollution	2.1064E-01 (7.4181E-02)	2.8396	59.8635 (31.8933)
Temperature Humidity Index	3.9523E-01 (2.9039E-01)	1.3610	54.1314 (8.2008)
Outpatient Medical Services			13.6919 (11.4311)
R^2	0.0864		

[a]The standard error of the regression coefficient is in parentheses. E-01, E 01, E 00, etc., indicate that the decimal place of the number is to be shifted to the left by one place; shifted to the right by one place; or not shifted at all.

[b]The standard deviation about the mean is in parentheses.

Appendix Table 3. S. I. P. S. regression results of statistical model 8. 2: respiratory diseases - lagged one day.

Variables	Coefficients[a]	t Statistic	Mean[b] - Untransformed
Sample Size	166		
Constant	-2.2162E 00	0.8446	
	(2.6239E 00)		
Age	-1.1754E-02	1.6309	44.9036
	(7.2070E-03)		(16.1028)
Sex	-2.2007E-04	0.1112	68.0723
	(1.9796E-03)		(52.8453)
Marital Status	7.2529E-04	0.2579	80.1205
	(2.8123E-03)		(40.0301)
Number of People in Household	-1.0366E-01	1.4000	3.3373
	(7.4044E-02)		(1.6933)
Household Income	-2.8430E-06	0.2011	11974.3976
	(1.4136E-05)		(7154.9744)
Race-Non-white	9.3073E-04	0.2309	6.6265
	(4.0313E-03)		(24.9497)
Drinking	3.1881E-04	0.1068	11.9235
	(2.9840E-03)		(35.8459)
Smoking Index	-2.1284E-04	0.5369	156.8643
	(3.9639E-04)		(260.8312)
Occupational Exposure Index	2.0080E-02	1.1286	3.8560
	(1.7792E-02)		(5.9076)
Air Pollution	2.0219E-01	1.4121	62.1369
	(1.4319E-01)		(34.5265)
Temperature Humidity Index	5.1959E-01	0.8170	52.8506
	(6.3601E-01)		(7.8279)
Per Capita Costs			1.7724
			(3.3421)
R^2	0.0603		

[a]The standard error of the regression coefficient is in parentheses. E-01, E 01, E 00, etc., indicate that the decimal place of the number is to be shifted to the left by one place; shifted to the right by one place; or not shifted at all, respectively.

[b]The standard deviation about the mean is in parentheses.