EVIDENCE FOR POSSIBLE ACUTE HEALTH EFFECTS OF AMBIENT AIR POLLUTION FROM TIME SERIES ANALYSIS: METHODOLOGICAL QUESTIONS AND SOME NEW RESULTS BASED ON NEW YORK CITY DAILY MORTALITY, 1963-1976*

HERBERT SCHIMMEL, Ph.D.
Associate Professor of Mathematics and Physics
Department of Neurology
Albert Einstein College of Medicine
Bronx, New York

OUR current studies examine the statistical association between daily levels of mortality and daily levels of air pollution in New York City to clarify and interpret this association. We are particularly concerned with the interplay between fundamental assumptions, techniques of analysis, and the statistical measures which are derived. The immediate focus of our studies is the 14-year period, 1963 to 1976, when approximately one and a quarter million deaths occurred in New York City. This period is of special interest because of a reduction in SO_2 levels such that the average SO_2 level of the first half of the period is approximately five times as high as that of the second half. One objective is to examine the influence of this reduction on measures of the mortality-pollution association. Statistical results relating to this question represent only a fraction of our objectives and are by no means the most significant results to be obtained.

In an immediately preceding study,[1] we emphasized the importance of proper consideration of the time-series structure of the data. One major objective is to demonstrate the importance of this, using the large data base available to us and to show how spurious results can be if adequate precautions are not taken.

*Presented as part of a *Symposium on Environmental Effects of Sulfur Oxides and Related Particulates* sponsored by the Subcommittee on Public Health Aspects of Energy of the Committee on Public Health of the New York Academy of Medicine and held at the Academy March 23 and 24, 1978.

Because of the length of this paper, it has been necessary to omit Appendix A and Tables A1 through A9 prepared by the author. Footnotes at appropriate points in the text refer to the omitted tables, which may be obtained by writing to the author. The first appendix appearing in this paper is Appendix B.

Address for reprint requests: Usonia Road, Pleasantville, N.Y. 10570.

This work on methodology tested against a large data base provides an improved methodology for similar time-series studies of both mortality and morbidity, and sets aside the results and conclusions of most earlier studies of mortality and morbidity, including some carried out at our institution.[2]

Our explorations are not yet completed. Moreover, as the analysis becomes more thorough, new difficulties are revealed (e.g., corrections for weather) that require further study.

One major difficulty in studies of the health effects of air pollution has been an implicit presumption of adverse effect which has permitted a looseness in instituting adequate safeguards in regression analyses and even greater looseness in interpreting results. The most glaring examples of this looseness with consequent error have been those studies which used SO_2 levels as the sole measure of air pollution. Setting aside the technical quality of such studies, almost always inadequate, a major impermissible error arises when measured health effects are attributed to SO_2 as a specific pollutant ignoring it as a general indicator of the level of air pollution.

The issue is joined when one considers the level of SO_2, i.e., the dosage. Certainly at high enough levels it can cause extremely adverse effects, but at ambient levels of 1 ppm. and below there is no conclusive evidence of either acute or chronic health damage, and at this level SO_2 can barely be detected by physiological response. In such cases, clearly analyses that relate levels of air pollution to levels of mortality or morbidity must control carefully for the effects of any extraneous variables, e. g., weather.

To appreciate our new results, both as to methodology and size of possible effects, it would be useful to present the background of our present study. This includes the strong evidence for adverse health effects during severe smog episodes, notably in London during the 1950s and 1960s. It also includes our studies that have circumscribed progressively the size and significance of possible effects.

Background: Evidence of London Episodes

Concern with "bad air" and sulfurous compounds in particular extends into antiquity.[1] During the past century a number of well-documented episodes have occurred where extraordinarily heavy smog was associated with immediate increases in mortality and morbidity. One such episode occurred in Donora, Pa., in 1948 and received intensive study, but no information was available on types of pollutants present.

The most dramatic episodes in recent history occurred in London after World War II, and the most severe lasted four days in December 1952. Smoke levels on the first day were about 20 times the average level and on the last two days were more than 30 times the average level in the winter. Mortality and morbidity began to rise within hours, and it was estimated that the episode resulted in approximately 4,000 to 5,000 excess deaths over a two-week period, more than double the expected number. Most of these occurred during and within a few days after the episode.

The authorities instituted smoke-abatement measures, and the recurring episodes finally disappeared in the mid-1960s. The most severe of the subsequent episodes in the 1960s occurred almost exactly 10 years after the first and resulted in approximately 700 excess deaths. The ratio of excess deaths of the 1962 episode to those of 1952 was approximately in proportion to the average smoke levels. Indeed, we have estimated that there was roughly one excess death for each 4 μg. of smoke per cubic meter per day during the two episodes.[1]

The London experience is of greatest importance for many reasons. Analysis of these data on smoke levels, mortality, and morbidity provide the best evidence to date that ambient air pollution, if high enough, has serious adverse health consequences. The London experience also demonstrated practical measures to cope with this clear-cut health problem, namely, smoke abatement. It is noteworthy that no specific pollutants were identified as being the harmful agents. Separate measurements of London SO_2 levels were available but separate controls were not adopted to reduce these. In fact, average SO_2 levels in both the 1952 and 1962 episodes were about the same (approximately 10 times the prevailing winter average) and reached daily averages of about 1.5 ppm. Although no specific SO_2 control measures were undertaken, the British students of these episodes suggested that SO_2 contributed to increased mortality and morbidity, but abandoned this hypothesis when recognizable dangerous episodes disappeared after sharp reduction of extremely high smoke levels without attendant reductions of SO_2 levels. Finally, analyses of the less severe episodes were confounded by the inherent Poissonian variability associated with daily mortality and the variability associated with weather changes.

Background Studies

At the time of the 1952 London smog episode, substantial interest in air-pollution control already existed in the United States and centered on

increasingly severe smog developing in the Los Angeles area and on smoke control and abatement generally.

Beginning with his appointment to our institution in 1960, Dr. Greenburg made systematic studies for 10 years of possible adverse health effects associated with New York City ambient air pollution.[2,3] Bearing in mind the London experience, these studies initially estimated excess mortality and morbidity associated with episodes of high pollution. These studies were inconclusive, although peak SO_2 values were approximately one half the 1.5 ppm. levels attained in the 1952 and 1962 London episodes.

Extending the episode approach, correlation and regression methods were used to compare mortality and pollution levels for long periods. In the last study of this series,[2] more than half a million death certificates were reviewed in an analysis of the mortality-pollution relation for the period 1963 to 1968. Our current work is essentially a lineal descendant of this work. In that earlier study 10 mortality classifications were employed. Correlation and partial correlation coefficients between daily mortality variables, two pollution variables (SO_2 and COHS or Smoke Shade), and seven weather variables were computed. Estimates of excess deaths were obtained by a regression of mortality on the two pollution variables, using three models: regression of daily crude mortality on the two crude pollution variables; regressions, using the above variables adjusted by eliminating annual and seasonal trends and same-day temperature and day-of-week effects; and regressions, where crude mortality and crude pollution variables were adjusted only for their relation to mean daily temperature. Regressions included same day and up to seven days previous effects. Among additional analyses included in the study were computation of cross-correlation functions to indicate precedence of changes in mortality, pollution and temperature variables, and a simplified theoretical and empirical treatment of the limitations on accuracy of estimates introduced by using a single measuring station for this large city.

Subsequent work has made clear that this study suffered from a number of major defects.

The term "excess deaths" was carried over from episodic analyses, although the term was circumscribed carefully for statistical purposes. This terminology reflected a presumption (i.e., bias) as to the existence of causal effects, a presumption that proved an important factor in selecting the wrong estimate as described below.

As will be explained more fully, the variables in these studies are a composite of annual, seasonal, and short-term, day-to-day changes. Because we can study only acute effects from these data, only day-to-day changes need be used for the analysis. This eliminates the problem of considering many factors in addition to air pollution that influence longer term, slower trends in the mortality variable. Method B, wherein annual and seasonal trends are eliminated by Fourier analysis from the variables before being introduced into a regression, proved a superior technique. This was not recognized and Method C, which gave estimates almost twice as large as Method B, was selected erroneously as providing best estimates.

Another major defect was failure adequately to take into account the influence on day-to-day mortality of previous days' temperatures. This was called to our attention by the publication by Buechley et al. of a study with similar objectives for approximately the same period for the New York City Metropolitan area.[4] We thought, and this later proved to be the case, that our failure to take into account the effects of previous temperatures would influence regressions where previous day pollution variables were introduced.*

Subsequent to the Schimmel-Greenburg study, air-pollution control sharply reduced ambient levels of SO_2 so that in 1972 they were approximately one sixth of the levels which had prevailed in the mid-1960s. We extended our data base and analyses through 1972 to determine how this reduction in SO_2 level influenced estimates of effects. Because of the major defects of the earlier study, we included a review and further development of methodology.

A new study by Schimmel and Murawski undertook to eliminate the major defects of the earlier one, and adopted Method B, with modifications. For technical reasons, a 15-day moving average was used instead of components computed by Fourier analysis to compute the slow rhythms to be eliminated from the variables. Previous day temperatures were introduced in the lag models.

New estimates were generally lower than those obtained in the previous study, and estimates of effects did not increase systematically as lag days

*The Buechley model employed only one measure of pollution, but included a variable for seasonal mortality trends and proved superior to our Method C, although, as explained later, we consider it inferior to the Method B approach we have continued to use. Analysis of residuals showed poor fit by pollution levels and by season. We believe that this is inherent in the model with a large number of control variables which interact with pollution variables. Moreover, the problem of accuracy of estimates based on a single station was exacerbated by including nearby urban and suburban areas. In a later study, Beuchley[8] included smoke and carbon monoxide measures in his model.

were introduced except for the respiratory disease classification. Separate analyses were carried out by three annual periods within the 10-year period and by two-month periods grouped for the entire 10-year period. The small size of effects relative to their estimated standard deviations precluded many of the interesting comparisons which were sought. The study found that the SO_2 variable tended to behave as an indicator or proxy for other variables when used as the sole measure of pollution, i.e., that as levels went down, regression coefficients went up. This behavior was not tested for statistical significance. When SO_2 and Smoke Shade were included in a joint regression, coefficients for SO_2 were usually not statistically significant although they tended to be weakly positive. Further, when regressions were carried out on particulate measure alone, the combined effect associated with SO_2 and Smoke Shade usually could be accounted for by the particulate measure alone.

Scope of Current Studies

Current studies largely respond to stated needs in the previous study. Specifically, new studies use an enlarged data base adding mortality, pollution, and weather variables for four additional years of low SO_2 levels to provide a more adequate basis to compare periods of high and low SO_2 levels. The enlarged data base provides more precise estimates for this and other purposes. Second, newer analyses are carried out for detailed age, sex, and race classifications. It was expected that this might increase precision of estimate, and serves the larger purpose of providing more homogeneous mortality classifications so that estimates are less affected by changing demographic and health conditions which, over time, alter the balance between these classes. Two additional causes of death classifications are introduced, each with its special significance: cancer and violence and accidents.

Current studies include more detailed studies of the mortality-temperature relation. Even after seasonal trends are removed, clearly daily mortality levels are influenced strongly by weather changes. Because pollution levels also are influenced strongly by weather changes, pollution-mortality relations cannot be determined without taking into account these other two mutual relations. This can be done easily if a simple same-day linear model is used, which, however, is not appropriate. The previous study showed that the relations were particularly complex if lag days are taken into account and, further, vary substantially from season to season. Efforts

in this area are necessarily limited because the study of mortality-weather relations could absorb readily all the resources available for the new studies.

The protocol for these studies also includes examination of the variation of estimates of effects as a function of the particular filters employed.

Finally, the protocol provides for further exploration of the limitations imposed by using a single station to measure pollution and the possibilities of using measurements from several stations. Although the air-pollution measurement network of New York City used as many as 40 stations during the early 1970s, collection of data has been irregular, especially the continuous daily series required for the analyses. The advantages to be gained by additional measurements are offset by the inaccuracies resulting from interpolating estimates for missing data.

Organization of the Present Report

Acquisition and validation of the extended data base have been completed only recently. In preparation, methodological questions have been theoretically and empirically explored, using the previous 10 years' data base.

In the next section, some of these methodologic questions are discussed at length. When we first became interested in the subject of this study we did not appreciate the complexity of the problem, and in this we were not alone. This led us, and we believe most others, to improper conclusions. Further, we find that the statistical tools that we employ are not understood by most epidemiologists and other environmentalists who are the essential audience for this work. This is partly due to failure to face up to the complexity of the problem, and partly to a lack of acquaintance by most statisticians of special approaches to time-series analyses. Throughout the discussion limitations are pointed out, some imposed by the very complexity of the problem and some by the available data.

The third section contains the statistical results available at the time of this report, including such results as may be available for Pittsburgh. In the final section these statistical results are discussed and their significance evaluated. An attempt has been made to document the methodology, for which appendices have been used to a considerable extent.

Review of Methodological Problems: Confronting the Complexity of the Problem

Questions Raised by the London Episodes

These episodes serve as a starting point for defining the complexity of the problem and the attempt to analyze it quantitatively. For the 1952 episode an average excess of about 1,500% in level of particulates for four days was associated with an excess of about 500% in mortality over these same four days. Actually, the excess accumulated over a period of about 10 days. If we test a simple linear hypothesis for the 1962 episode and for the somewhat more severe episode of 1956 we find a crude consistency. Simple calculation yields an estimate of 30% to 35% excess deaths associated with the prevailing nonepisodic winter levels, i.e., when we extrapolate downward to zero pollution. Before publication of the Schimmel-Greenburg study, such a calculation and a comparison with other published studies were made, and it was this calculation as much as any other factor which led to the exaggerated estimates under Method C of that study. No valid study is available to test this linear hypothesis for London episodes, and our further study of New York City suggests rejection.

The British analysts suggested threshold levels below which no adverse health effects were measured by the association with increased mortality. Fortunately for the people of greater London, severe episodes diminished and disappeared before providing adequate data for analysis. Morbidity as well as mortality were studied for a number of the London episodes, and morbidity did increase but mortality proved as sensitive a measure of the pollution effect as morbidity. Further, it was easier to define the mortality variable, and mortality data were more readily available. Consequently, health damage for the London episodes has been measured principally by excess mortality.

Is daily mortality a well-defined variable? It is a well-defined number measured relatively accurately for London and New York City. There may be minor inaccuracies. For example, Dr. Leonard Greenburg suggested in a personal communication that some Sunday deaths are filed as Monday deaths on death certificates, which contributes to higher Sunday than Monday averages, but this is a minor systematic error. Certainly as a statistical variable, total daily mortality for London, New York City, and many other areas is about as well defined quantitatively as the studies

require. It is certainly better defined than virtually any available city-wide morbidity measures. Recognizing that the mortality variable is accurate does not overcome a fundamental difficulty in using it as a measure of the adverse health effect of air pollution, be it limited as it must in this case to the acute effect in case of episodes. Autopsies for the London episodes, as for the earlier Donora episode, did not identify any site for or mechanism by which the ambient air pollution could have caused death. This means that even though these London episodes meet certain criteria of a simple deterministic causal model, including the disappearance of effect with the removal of cause, we have essentially a statistical finding. Lacking detail on mechanism, we could still give substantial meaning to the mortality measure of adverse health effects if we had adequate statistics to show, in addition to the number of excess deaths, the reduction in expected life span for those who died. More specifically, are these excess deaths accounted for by terminally ill individuals who would have died within a few days or weeks, or do these result from exposure to unusual stresses of comparatively healthy people who otherwise would have lived many years?

The question is of considerable importance. If, for example, we were to conclude that the pollution-mortality relation for New York City is causal and that there is an average excess of 3% deaths for the average pollution level, it might seem to be a very strong effect. Yet, if the prematurity of death were only a week and air pollution were reduced gradually from the current level to zero within one year, the expected reduction in annual mortality would be only 0.03% for the first year and none thereafter. The expected Poissonian variability is approximately 0.3%. On the other hand, if the expected prematurity averaged six months, the first year's expected reduction would be about 0.6% and comparatively small reductions thereafter for a total of about 0.15% mortality. Unfortunately, there is no detailed information on prematurity as has been pointed out in the previous study,[1] although there are some clues. Thus, for the more severe 1952 London episode, the group above and below 70 years of age showed approximately an equal response as measured by an increase in mortality, but in the far less severe episodes of 1962 the older group showed about twice the response of the younger group. One can theorize that air pollution represents a stress factor that increases with the level of air pollution. At very low levels the death of a terminally ill patient may be hastened by hours or days. At high levels an acutely ill individual otherwise likely to survive may succumb to the added stress. This is a reasonable hypothesis,

but one for which we have only suggestive data. Classification of mortality by cause of death and age throw some light on mortality as a measure of adverse health effects, but thus far have provided no significant quantitative information on the extent to which life is shortened.

Cigarette Smoking as a Form of Air Pollution

One may ask whether chronic and permanent damage is associated with ambient air pollution. Just as the severe London smog episodes suggest that acute health effects should not be excluded at lower levels, studies of cigarette smoking might provide useful insight into this question, particularly in view of the wealth of detail contained in studies of smoking. Of course, the mix of pollutants differs from ambient air pollution, and, when smoking, an individual is exposed to levels of pollution for some substances an order of magnitude greater than prevailing city air, and some of these levels are, in fact, 100 to 1,000 times greater than accepted environmental and occupational standards. For our purposes, we wish to distinguish between acute (or temporary) effect and permanent effect. The former classification would include such effects as heart attacks and strokes which diminish sharply if the pollution (cigarette smoking) ceases and those such as lung cancer and emphysema, where the effect remains after withdrawal. This discussion is in terms of risk, and the effect is measured statistically. It has been established that for regular smokers of one pack a day and more, a risk for heart attack and stroke which is more than twice that of non-smokers is reduced to almost the same level as that of the non-smokers when smoking ceases. In terms of numbers of individuals affected by morbidity and mortality, these two classifications account for the largest part of the adverse health effects. This finding suggests that studies directed toward determining the acute effects of ambient air pollution should be able to estimate the major part of the effect associated with chronic exposure as it relates to these two classifications of morbidity and mortality. It also suggests that studies of permanent damages associated with differential chronic exposure to pollution should focus on differential rates of emphysema, lung cancer, and less severe permanent or chronic respiratory disease.

Treating cigarette smoking as a form of air pollution can give some insight into the complexity of examining acute health effects of exposure to low levels of SO_2. A nonsmoker detects the presence of cigarette smoke at lower levels than a smoker, and is irritated by cigarette smoke at lower

levels than regular cigarette smokers. There are operating here processes of loss of sensitivity, adaptation, and habituation. While sensitivity may decline, frequently sufficient sensitivity remains for even a heavy smoker to suffer continuous irritation, as indicated by coughing and irritation of the eyes.

Turning to SO_2, and without attempting to be precise in every detail, detection of the presence of SO_2 is reported to occur in the range of 0.7 ppm. to 2.0 ppm. with considerable individual variation. Detection occurs usually as mild irritation of the eyes, and irritation increases with dosage. Tolerable levels have been reported by workers at pulp and paper mills up to 100 times higher. A pollutant can be acutely damaging, e.g., carbon monoxide, without conscious detection, but its presence is detected by a physiological effect. The range of 0.7-2.0 ppm. approximately corresponds to the threshold for the principal physiological effect in humans which indicates the presence of SO_2, namely, increased airway resistance. Here, individual response varies even more, but the response is within tolerable physiological limits. Moreover, adaptation is rapid and the response disappears on continued presentation of the SO_2 stimulus. This is important to interpretation of the results of our regression analyses when compared with literature on human and animal experiments and occupational experience. Yet, if the level of ambient air pollution influences mortality, it seems to us that it does so primarily by increasing the risk for those who are already at very high risk of death. The experiments just discussed were not carried out on such individuals, e.g., on individuals in intensive care units or terminally ill.

Having surveyed the relevant literature, we suggest that the most important studies which could throw light on the possible acute effect of ambient air pollution have not been carried out, namely, exposure to controlled environments of animals already at high mortality risk, e.g., 10% to 20% per week.

The Causality Question: Hypothesis Versus Demonstration

Why a causality hypothesis? The statistical analyses carried out in this study are directed toward a public health problem of immediate public concern. This is the purpose of exploring the mortality-pollution relation, a relation which in the presence of many other factors influences mortality. Some of these factors also may influence pollution or even may be influenced by pollution. It is thus essential to recognize the context in

TABLE I. VARIATION OF "EFFECTS" EXPRESSED AS % MEAN DAILY MORTALITY ASSOCIATED WITH SO_2 AND SMOKE SHADE POLLUTION WITH METHOD OF ANALYSIS*

Regression method	*Period*	*SO_2*	*Smoke Shade*	*F**
1	1963-1968	1.88(*)	9.80(*)	263.0
2	1963-1968	1.50(*)	6.11(*)	67.4
3	1963-1968	1.26(*)	2.85(*)	12.6
4	1963-1972	0.58(0.51)	2.20(0.79)	
5	1963-1972	-0.18(0.63)	2.12(0.89)	
6	1963-1972	-0.25(0.44)	2.04(0.61)	
7	1963-1972	-0.25(0.37)	2.04(0.53)	
8	1963-1976	0.20(0.27)	1.39(0.42)	
9	1963-1976	0.20(0.28)	0.95(0.42)	
10	1963-1976	-0.35(0.26)	0.94(0.40)	
11	1963-1969	-0.23(0.43)	1.23(0.61)	
12	1970-1976	-0.43(0.26)	0.69(0.53)	

Standard Error of Estimated Effects in Parentheses

The measure of "effects" is the Mean Standardized Regression Coefficient as defined in the present study and corresponds to "% Excess Deaths," as used in the Schimmel-Greenburg (S-G) study and "% Premature Deaths" in the Schimmel-Murawski (S-M) study. It is clear from the present study and from the earlier S-M study that the large variation in estimates is due primarily to the method of analysis, rather than to the years covered by a particular analysis.

All coefficients are derived in a regression of Daily Total Mortality (M_1) jointly on same day Mean Daily SO_2 (P_1) and Smoke Shade (P_2). Particular estimates vary according to processing of these variables prior to regression and the inclusion of temperature variables in the regressions as described below:

1,2,3. These estimates represent recomputations from the data reported for the S-G study, and represent Methods A, C, and B, respectively. Under 1, namely A, crude M_1 was regressed on crude P_1 and P_2. This estimate includes no control for temperature or season. Under 2, namely C, crude M_1 was regressed on P_1 and P_2 variables which had been regressed on a quartic of crude mean daily temperature (T). This method apparently corrected in part for seasonal relations and in part for local (in time) temperature effects, notably heat waves. It was erroneously accepted as the preferred best estimate. Under 3, P_1 and P_2 were measured as deviations of crude P_1 and P_2 from trend lines defined by a Fourier analysis. They were further corrected for same-day relations with a deviation daily temperature computed as for P_1 and P_2. This method corrected for the interactions of slow components, but not to the extent the filtering method used in the present report.

4. These estimates were obtained in the S-M study, where crude M_1 was regressed on filtered same days P_1, P_2, and T. Filtered variables were defined as deviations from a 15-day moving average.

5,6,7. These estimates were obtained in the course of some exploratory studies in which deviation temperature with some lags was used to correct deviation P_1 and P_2. 5 employed crude M_1 as the dependent variables, and 6 and 7 employed, respectively, deviation mortality and deviation mortality corrected for temperature in the same way as P_1 and P_2. This was essentially a test of the use of the hybrid model of the S-M study where crude mortality was regressed on filtered P_1, P_2. The test showed that the hybrid model gave satisfactory estimates in same-day regressions. Further tests showed that the hybrid model was inadequate for lags with regressions.

8. In this regression deviation M_1 is regressed on deviation P_1, P_2 and T. See further discussion below.

9, 10. The estimates in 10 were obtained using deviation M_1, P_1 and P_2 corrected for temperature as described in Appendix D. 9 are estimates obtained when the temperature variables used for the correction were introduced directly into a regression using deviation M_1, P_1, and P_2. Estimates would be identical if temperature corrections had been made on M_1, P_1, and P_2 using pooled data for the period. However, as described in Appendix D, data were corrected by pooled months.

11, 12. These are estimates obtained in the same way as 10. There are differences between the early period and the later period. These are not statistically significant, but correspond to the reduction in mean values of Smoke Shade between the two periods.

which results may be used and to attempt to explore the various possibilities of measuring the statistical association with this in mind.

Frequent use of the words "effect" or "dose-response" relation will, therefore, not imply demonstration of causality. A high level of statistical significance for a particular measure of association, using a null hypothesis, should not, of itself, show a causally demonstrated relation. As will be seen, measures which from a causal approach are most objectionable show the strongest effects at the highest levels of statistical significance (see Table I), but failure to reject the null hypothesis does not of itself demonstrate absence of causal effect. It is only useful in rejecting particular hypothetical models and in excluding strong causal relations in accordance with these models, e.g., three times the standard error of estimate of the effect. However, it may well be that an appropriate model would show statistically significant effects. For example, if a simple linear model is employed where a threshold model is appropriate and if this threshold is exceeded by only 5% of the 5,000 sample days, this effect could be lost in the pooled data. For example, if effect is measured by correlation coefficient, the coefficient would be approximately 2.5 times higher for a threshold model.

Clarifying causal relations. To attempt to clarify the meaning of causality in a profound sense would involve philosophical considerations which would be diversionary to the immediate study. Causality is here used in a pragmatic, even scientifically vulgar, sense. In the real world, causal quantitative relations are not established easily, even in the physical sciences where progress takes place over centuries.

In the health sciences, and particularly in epidemiology, where intervention is limited, many factors may be present which need to be disentangled. In an excellent treatise with many illustrations of the complexity involved, Susser[7] introduces the terms, "Hypothetical Causal Variables", "Control Variables," "Uncontrolled Variables," "Confounding Variables," "Distorter Variables," and "Suppressor Variables." This is done in the chapter[8] dealing with Screening Causal Models for Extraneous Factors.

Strong effects versus weak effects. In the task at hand, it is useful to distinguish between strong effects and weak effects. In controlled experiments, e.g., drug studies, a portion of the variation in response will be associated with uncontrolled variables and a portion with control variables, e.g., age or sex. The stronger the effect (i.e., response), the more easily it

will be detected and measured with some degree of precision. In general, such situations are governed by a statistical law which requires that the number of samples (i.e., patients, cases, subjects) needed increases as the inverse square of the effect. Thus, if the effect is reduced by one half, the number of samples must be increased fourfold.

The situation is far more difficult and complex than indicated by this general observation. In any analysis, the influence of extraneous variables of comparable or greater strength than the hypothetical cause must be considered with some degree of precision. The greater the influence of extraneous variables, the greater the degree of precision required in estimating their effect. As the size of the effect of the variable or variables of primary interest is reduced, the influence of previously incorporated extraneous variables must be measured with increasing precision. Further, additional variables of relevant influence must be introduced into the analysis.

The severe London smog episode of 1952 was analyzed with only perfunctory attention to weather variables, and when attempts were made to analyze later episodes, particularly the less severe ones, it became clear that weather factors had to be taken into account. Consider the London experience as primitive experiments and studies with strong effects out of which came the conclusion that particulates were the primary causal agent and smoke abatement measures were taken accordingly. The ensuing experience tended to confirm the original conclusion.

In contrast, New York City had no severe smog episodes. During the 1960s public measures concentrated on reducing incineration whose gross particulates were never demonstrated to be adverse to public health but were clearly an aesthetic nuisance. Beginning in 1969, public measures focused on reducing the level of SO_2 in the ambient air by 85% by 1972 while simultaneously particulates were reduced by a relatively moderate 25%. First attempts to evaluate statistically the effect in reduction of SO_2 levels showed that SO_2 behaved as an indicator or proxy rather than as a harmful pollutant.[1-4] It is interesting to speculate what would have been the London experience had public measures been directed primarily toward reducing SO_2 instead of particulates and what the New York experience would have been if public measures had been directed toward reducing respirable particulates.

In any event, we are dealing with relatively small effects requiring large data bases and careful analysis of extraneous variables. In this connection

it is useful to contrast the relative strength of weather effects with those of air pollution. Data are available for estimated coefficients and their values obtained by regressing 56 daily mortality variables on the daily perturbations (i.e., deviations from a moving average) of mean daily SO_2, Smoke Shade, and temperature.* For this purpose the 1963 to 1972 data base of the previous study was employed, using the regression methods of that study. The 56 classifications were defined by seven cause-of-death classifications broken down into eight classifications by race, sex, and age. Since there were only six independent death classifications and six independent race, sex, and age classifications, only 36 of the 56 classifications were independent. Using conventional criteria which, however, are not entirely appropriate because there is some sequential correlation, the table shows that 29 of the 56 temperature coefficients, all positive, are significant at the $p < 0.01$ level. In contrast, none of the coefficients for SO_2 is significant at this level. Of the three SO_2 coefficients significant at the $p < 0.1$ level, one is negative. The Smoke Shade coefficients occupy an intermediate position. That temperature is a stronger factor than air pollution need not of itself confound our analysis except that perturbations in temperature correlate significantly with perturbations in pollution. This is understandable, because levels of these pollutants, produced primarily by stationary sources, are influenced largely by the daily production level and daily weather pattern. This collinearity of variables means that estimates of pollution effects can be influenced substantially by the manner in which weather variables are introduced in the regression. This question receives greater attention in our present study than in our earlier studies.

Before leaving the question of causality, it is well to consider a hypothesis put forward by Mervyn Susser:[6] that some of the effects associated with weather changes should be attributed to pollution because changing levels of air pollution may cause weather changes. Certainly this possibility cannot be dismissed in terms of the physical mechanisms available, e.g., screening of light, catalysing fog, and precipitation. However, day-to-day changes such as those encountered in the temperature variable used in the present analyses appear to provide no evidence for significant influence.

Weak effects and a built-in bias in statistical studies. When hypothetical effects are weak and the subject is of sufficient scientific or public interest

*These data are available from the author as his Table A1.

to encourage numerous studies, a built-in bias favors reporting of detection of effects because authors usually do not report all the negative analyses which have been tried. This situation may prevail in the environmental field at the present time. Accretion of such studies need not be evidence of causality in terms of such usual criteria of causality as consistency and replication. Each study must, itself, stand up to critical examination. The analysis carried out here and in similar studies should be considered as suggestive rather than demonstrative of causality, and corroborating evidence should be sought from other areas, notably human and animal experiments and occupational health studies.

Alternative Regression Approaches to Estimating the Pollution-Mortality Relations

Studies carried out by Buechley[4,8] introduce crude mortality as the dependent variable, pollution variables as independent variables, and a number of additional independent (or control variables) including temperature. The object of the analysis is to account for as much of the variability as is possible and to allocate that portion which may be properly assigned to the pollution variables. This is not easy because pollution variables correlate with other variables and the proportion of variance accounted for by pollution variables depends on the sequence in which variables are introduced. The analyses assume a linear relation between the daily pollution variable and its effect on daily mortality. The Buechley approach is essentially the one adopted by most investigators.

The approach first introduced in Method B of the Schimmel-Greenburg study eliminates the annual and seasonal trends for pollution variables, and regression coefficients measure only the response of the mortality variable to the high-frequency fluctuations of pollution. If one adopts the linear hypothesis, regression coefficients that describe the linear relation are independent of the frequency of fluctuations of the pollution variable, particularly if the analysis is limited to same-day regression. If lag effects over a few days are also to be estimated, then the independent pollution variable should contain sufficient separate frequencies to estimate separate regression coefficients for the lag days. Why adopt this estimation of the regression coefficients from the high-frequency components? On the hypothesis that mortality increases monotonically with increased pollution, daily variation of mortality will respond to daily changes of the pollution

variables and one need include in the analysis only those variables which are jointly correlated with pollution and mortality at high frequency. Only weather variables and day of week have thus far been identified as fulfilling this role. When annual and seasonal behavior are included in the independent pollution variables, the estimated regression coefficients measure the response to the variation present in the annual and seasonal behavior of the mortality variable. To allocate the proper effect to the pollution variables, it is essential to include all the additional variables which may influence the annual and seasonal pattern. To the extent that such variables are not properly introduced in phase and shape, they cause error in estimates of coefficients for pollution. Omission of relevant variables also causes error.

Development of Filtered and Temperature-Corrected Variables for Use in Regression Analysis

Mathematical Statement of the Approach

It is assumed that the presence of pollution causes acute additive effect ΔM on daily mortality which can be expressed as a linear form:

$$\text{Equation 1.} \qquad \Delta M = \sum a_k Q_k$$

where the Q_k are functions of the daily pollution variables P_j. They may be nonlinear forms, as, for example, squares and higher powers and cross products; or they may be the positive excess above a threshold level; or they may be the positive excess over yesterday's level if one considers that only increases have an adverse effect because of adaptation. The Q_k can take into account interaction of pollution variables with other variables. For example, if one hypothesizes that SO_2 has an adverse effect if sufficient humidity is present, then a Q_k variable can be introduced of the form HP_1 where P_1 is the SO_2 level and H is a function of humidity. An even more complex Q_k can be defined and tested, as, for example, a function of P_1, humidity, and temperature.

An average effect over the period of analysis will be defined by

$$\text{Equation 2.} \qquad \Delta M = \sum a_k \overline{Q}_k$$

where $\overline{Q}_k$ is the average of the Q_k for the period. As has already been pointed out, this definition is ambiguous if it is assumed, as one must, that

the individuals represented as ΔM would have eventually died. Then, if sufficient lags are taken into account ΔM must always be zero and the parameters which would be of interest would be those that describe the prematurity. For example, even a substantial portion of deaths a few days early would not reduce the annual daily average of mortality. During the present study, an attempt will be made to take into account lags up to two weeks. However, because of increasing error of estimates it is unlikely that significant estimates of prematurity will be obtained. If the Q_k is normalized so that $\overline{Q}_k = 1$ and M is normalized so that $M = 100$, then ΔM is a percent and given by:

$$\text{Equation 3.} \qquad \overline{\Delta M} = \sum a_k;\ \overline{M} = 100;\ Q_k = 1 \quad .$$

To estimate the a_k a linear least squares regression model will be employed of the form.

$$\text{Equation 4.} \qquad M = (\Delta M - \overline{\Delta M}) + F(X) + \epsilon$$

where F(X) is a function with linear coefficients of functions of all extraneous variables which explain mortality, and ϵ is a residual term (i.e., error term) which includes Poissonian variability. Note that the omission of extraneous variables which correlate with the Q_k will cause errors in estimates of the a_k. Variables which influence M but are not correlated with the Q_k may be omitted with some reduction in accuracy of estimates.

It should be noted that a constant added to any of the Q_k would yield the same estimates of regression coefficients but would alter the value of Q_k because the new mean would be the old mean plus the constant. In the case of the problem at hand and particularly for the simplified linear model which will be adopted, there is a natural zero so that Q_k becomes well defined. Moreover, this natural zero for $P_1 = SO_2$ and $P_2 =$ Smoke Shade is virtually within the range of the variables because sample days occur when pollution is close to zero.

Simplified Linear Model for Effect of Pollution

For the present report, it is assumed that:

$$\text{Equation 5.} \qquad \Delta M = \sum a_{1,j}P_{1,j} + \sum a_{2,j}P_{2,j}$$

where $P_{1,j}$ is the SO_2 value for the jth day preceding the mortality variable and similarly $P_{2,j}$ is the Smoke Shade value. The combined average effect in percent is:

Equation 6. $$\overline{\Delta M} = \sum a_{1,j} + \sum a_{2,j};$$
$$P_1 = \overline{P_2} = 1; \overline{M} = 100$$

We can associate $\sum a_{1,j}$ and $\sum a_{2,j}$ as the partitioned effects, respectively, for SO_2 and Smoke Shade.

The actual model need not be the simple linear model given by Equation 5, but can include the influence of slowly changing extraneous variables which result in the $a_{1,j}$ and $a_{2,j}$ possessing a seasonal pattern. To check this, the regressions should be carried out by seasons or months.

Application of Filtering to Simplified Model

If we apply a linear filter to both sides of Equation 4, the relation of equality will apply to the filtered variables on each side. The advantage of such a filter is that if we pass only fast, i.e., high-frequency components we shall filter out all extraneous variables which have slow components. All such variables which influence the slower movements of the mortality variable would tend to correlate with the pollution variables and would, therefore, have to be included in the regression equation if filtering were not employed. We assume that after filtering the regression equation will now have the form:

Equation 7. $$M' = \sum a_{1,j} P'_{1,j} + \sum a_{2,j} P'_{2,j} + F'(T) + \epsilon'$$

where the prime denotes the fast component and $F'(T)$ is a linear form of temperature functions. The temperature functions themselves are not assumed to be linear. The linear filter employed for the present report yields the deviation of a variable from its 15-day centered moving average so that the variable is decomposed into a slow component, the 15-day moving average, and a fast component which is the deviation from the moving average. The filter attenuates rhythms of periods over 15 days so that the amplitudes of cosines of periods 30, 75, and 150 days are reduced to 30%, 10%, and 2% of their prefilter values, respectively. Thus, the influence of seasonal movements is reduced substantially.

Elimination of Annual Trends

Before applying the filtering operation, each of the mortality and pollution variables is divided by a 365-day centered moving average. As a result, these variables tend to have an average value of one over any

365-day period or multiple thereof. The new variables, denoted as "indexes" or "index form," retain seasonal and faster movements. For the regression, this index form of mortality is further multiplied by 100, so that it has a mean of 100 over a period of years. In the case of the mortality variable, the standardization assumes that a fixed pollution effect results in a percentage increase in mortality. For the pollution as well as mortality variables, the index form assumes that a given temperature function affects these variables proportionately. This assumption seems reasonable.

The index form also is appropriate for the pollution-mortality relation if the pollution variable serves primarily as an indicator of day-to-day changes of effective pollution independent of the slowly changing absolute level. The index form is not appropriate if the effects are proportionate to absolute levels. However, if the annual levels are changing slowly it should give a fairly precise estimate corresponding to the mean level for the period. The index behavior can then be tested by comparing two periods with substantially different means. If it is behaving as an index, i.e., surrogate or proxy for other pollutants whose absolute values remained stationary, the estimated effect should be the same for the two periods regardless of the difference in absolute means for the proxy variable. If, however, the effect is proportionate to the absolute level then the percentage effects estimated for each period should be proportional to the absolute means of the period. In this way index behavior can be tested.

Adjustments for Outliers Prior to Filtering

Because least squares methods are particularly sensitive to extreme values, all mortality and pollution variables outside of the range of three standard deviations about the mean were given values of plus or minus three-standard deviations as appropriate. The details of this adjustment are described in Appendix B. Fewer than 2% of sample values for any variables were so adjusted.

Adjustment of Pollution Variables for Missing Data

Data have been assembled from seven monitoring stations for SO_2 and smoke levels use in regression analyses in accordance with the note in Appendix E as part of the complete study. For the time being, only a single series of SO_2 and Smoke Shade values are employed in regression

analyses. For the period 1963 through the middle of 1975 these are from the laboratory station, and for the balance of the period through 1976 this series has been extended in index deviation form, using data from the Roosevelt Island station. Data from this latter station and from other stations were also used to interpolate data missing from the laboratory station series before mid-1975, as explained in Appendix C.

Temperature Variables

In the earlier Schimmel-Murawski study, F(T) was defined as mean daily temperature and the j values had the same range as the $P_{1,j}$ and $P_{2,j}$ variables. As discussed earlier,* in a joint regression of 56 mortality variables on the fast components of same-day SO_2, Smoke Shade, and mean temperature, more statistically significant associations with temperature were found than with pollution. Because of the strong effects of temperature, it was decided to introduce more extensive functions of temperature in defining F(T). For the present report, F(T) includes same day and lag temperatures, separate variables for positive and negative deviations from expected temperatures for the same day and for the average of the same and previous day, the squares of these latter deviations, and a special deviation variable for the May through August period. The exact definition is shown in Appendix D.

The linear form F′(T), the filtered or fast component of F(T), was not introduced directly into the regression equation. Instead, it was used to correct the pollution variables and form new variables $P'_1{}^*$ and $P'_2{}^*$. It can be shown that if such F′(t) correction is applied simultaneously to the sample days of the whole period that F′(t) can be omitted and ΔM defined as

$$\text{Equation 8.} \qquad \Delta M' = \sum a_{1,j} P'_{1,j}{}^* + \sum a_{2,j} P'_{2,j}{}^*$$

will yield the same estimates for the regression coefficients as Equation 7. Moreover, if ΔM′ is also corrected by F′(T) the standard error of estimate will be the same; if not, it will be somewhat larger.

In the previous study it was observed that the regression coefficients for the temperature functions of a simpler F′(t) varied widely with the season of the year. Accordingly, in the present study F′(t) was used to correct the variables separately for each month with samples pooled by years (see Appendix D).

*These data are available from the author as Table A1.

TABLE II

		Mortality classifications	
		1963-1968	*1969-1972*
M_1	All causes	1-999*	1-999*
M_2	Respiratory	1-8, 470-527	10-12, 460-519
M_3	Heart disease	410-416, 420-434	393-398, 410-429
M_4	Other circulatory	330-334, 400-402 435-468	390-392, 399-409, 430-458
M_5	Cancer	140-239	140-239
M_6	Accidents, violence, etc.	800-999	800-999
M_7	Other		
$M_{10} = M_2 + M_3 + M_4 + M_5$			

Sex, Race, and Age classifications

M—Male, F—Female, W—White, O—Other
Age by years as shown. Initial breakdowns by age included 1, 1-59, 60-74, 75+, all. Combined categories indicated by—as M+F or W+O.

*International Classification of Disease Groups

In Table I, lines five, six, and seven suggest that estimates vary little if the M form or M′* form is used as the dependent variable. Prime (′) denotes fast component and star (*) denotes temperature correction. This appears to be the case for same-day regressions but empirical tests show that this is not so when lag variables with a range $0 \leq j \leq 7$ are used because of crossover frequencies where the slow component of mortality interacts with the pollution variables from which these slower components have not been removed entirely. As lag values are extended, the slower frequencies play a more prominent part in the estimates. Table I also shows in lines nine and 10 the influence on estimates of introducing F′(T) directly into the regression as compared with using it to correct the mortality and pollution variables on a month-by-month basis.

FURTHER QUESTIONS AND POSSIBLE SOLUTIONS

Mortality variables and effects. Extensive demographic changes have occurred over the 14-year period of analysis. The question has been raised whether these changes cannot, of themselves, introduce changes in estimates of effects when the early and later periods are compared. The creation of homogeneous classifications by race, sex, and age as well as by cause not only provides additional insight into the mortality-pollution

relation but also provides an adjustment for these demographic changes. A large number of classes have been defined as shown in Table II. Only the larger classes will be analyzed in detail.

The absence of any quantitative estimate of prematurity of death is a basic deficiency of the measures developed here. Extension of the lag regressions to take into account pollution levels as much as two weeks preceding death is unlikely to be helpful because of the large sampling error of estimate. Use of a simplified model containing as few as four parameters to describe the dose-response curve could provide the necessary precision of estimate, and it is planned to test this as part of the completed study. Finally, although we consider that none of the data developed in such statistical studies can be considered as more than suggestive of causality, development of measures which take into account regression coefficients where pollution lags mortality should show the relevance of the estimates to the causal hypothesis. Such analyses are included in the results section.

Problems connected with pollution variables. A number of problems are connected with the use and interpretation of results as far as these variables are concerned. These include the accuracy of instrumental measurements, what is being measured, the significance and usefulness of local station measurements, and the allocation of measured effects.

As to accuracy of measurements, the analysis can tolerate a substantial amount of random error in individual measurements because more than 30,000 Smoke Shade and more than 60,000 SO_2 measurements are employed for each seven-year period pollution series. Even systematic bias in error should not influence results substantially provided it is invariant over the period of study.

What is being measured confronts us with an anomaly. On the one hand, the SO_2 measurement presumes to be a measurement of a very specific substance although it has been suggested that some instrumental methods include other gaseous oxides. On the other hand, the Smoke Shade measure, whether calculated from transmittance or reflectance, is a crude measure of small particulates, provides no information as to particulate composition, and is sensitive not only to the level (amount) of particulates but also to their changing composition and size. With changes in types of fuels which have occurred over the 14-year period, it is unlikely that it can be considered a stable measure.

It is well known that local measurements vary from nearby and even

side-by-side locations, and a correlation of only 0.7 between two measuring instruments is not unusual in the latter case. When, in addition, only one or a few stations are used for pollution measurements the question arises as to how useful pollution measures are for analysis. For the present report and for earlier studies, measurements from a single station, the laboratory station, have been the principal data used as pollution variables: How representative are these measurements of the city-wide situation? A study of the New York City air-monitoring network by Dr. Inge Goldstein showed that the average correlation coefficients between the laboratory-station measurements and those of almost 40 other stations was about 0.5. Our own analyses of correlations with the Roosevelt Island station showed a correlation for fast components of 0.7 for SO_2 and 0.6 for Smoke Shade. Both of these stations, and particularly the laboratory station, are located in the area of highest level of pollution.

The question arises also with respect to annual trends as to how representative the annual trends at the laboratory station are for the city as a whole. In a special report covering the period 1957 through mid-1976, Dr. Ellis Blade, formerly of the New York City Department of Air Resources, plotted annual trends for a number of stations available from time to time during this period and visual inspection of semilog plots shows good agreement in annual trend data.

Analysis of a hypothetical ideal model, described in Appendix E, indicates that the use of a single station or even several stations would tend to give estimates below those obtained by an ideal model. The degree of underestimation would depend on how representative the single or several stations were of local conditions. It should be added that even though estimates of combined effects tend to be low, if the pattern is stable, valid comparisons can be made between estimates for two periods.

It should be noted that the hypothetical ideal model is in no way realistic. Movement of people means that local measurements are meaningless for much of the populations, and local variation between indoors and outdoors is extremely high.

Turning now to the allocation of measured effects, most of the ambient pollutants produced by stationary combustion sources tend to be correlated. Thus, SO_2 and Smoke Shade variables have correlation coefficients of about 0.6 at the laboratory station. Differential time series patterns can arise in two ways: by different rates of activity at local sources with different pollutants and by different rates of diffusion, convection, and

TABLE III. MEANS AND STANDARD DEVIATIONS (IN PARENTHESES) OF PRINCIPAL DAILY VARIABLES, CRUDE FORM, BY PERIODS

Period	M_1	M_{10}	P_1	P_2	T
1963-1969	243.1	190.0	16.3	21.5	54.2
1963-1969	(29.8)	(25.7)	(10.4)	(10.1)	(17.5)
1970-1976	225.2	174.2	4.2	17.7	54.7
1970-1976	(28.4)	(23.7)	(2.5)	(8.7)	(17.5)
1963-1976	234.1	182.1	10.3	19.6	54.5
1963-1976	(30.4)	(25.9)	(9.8)	(9.6)	(17.5)

M_1 = total mortality; M_{10} = adjusted total mortality (see Table II), P_1 = mean daily SO_2 in parts/million x 100, P_2 = mean daily Smoke Shade in COHS x 10

TABLE IV. CORRELATION MATRIX FOR ROOSEVELT ISLAND AND LABORATORY-STATION POLLUTION VARIABLES; COEFFICIENTS FOR INDEX-FORM VARIABLES *UNDERLINED* AND FOR FILTERED INDEX VARIABLES IN PARENTHESES. JANUARY 1971 to MAY 1975

	Lab station		*Roosevelt Island*	
	SO_2	*Smoke*	*SO_2*	*Smoke*
Lab station SO_2	1.00	*0.54*	*0.63*	*0.55*
Lab station smoke	(0.49)	1.00	*0.44*	*0.60*
Roosevelt Is. SO_2	(0.70)	(0.39)	1.00	*0.51*
Roosevelt Is. smoke	(0.58)	(0.59)	(0.55)	1.00

other transport. The latter can be influenced strongly by local terrain and smokestack heights.

For the New York City area, the two pollution measures used not only measure their own effects but such other pollutants as may be correlated with them.

Statistical Results

SCOPE OF RESULTS — DATA FILE

The data file has been expanded in a number of directions since the previous study. Death-certificate information has been obtained for four more years. The mortality file contains details on approximately one and a quarter million deaths that occurred in New York City during the period 1963 to 1976.[2] Daily mortality variables are computed for the classifications listed in Table II. For the present report only analyses based on

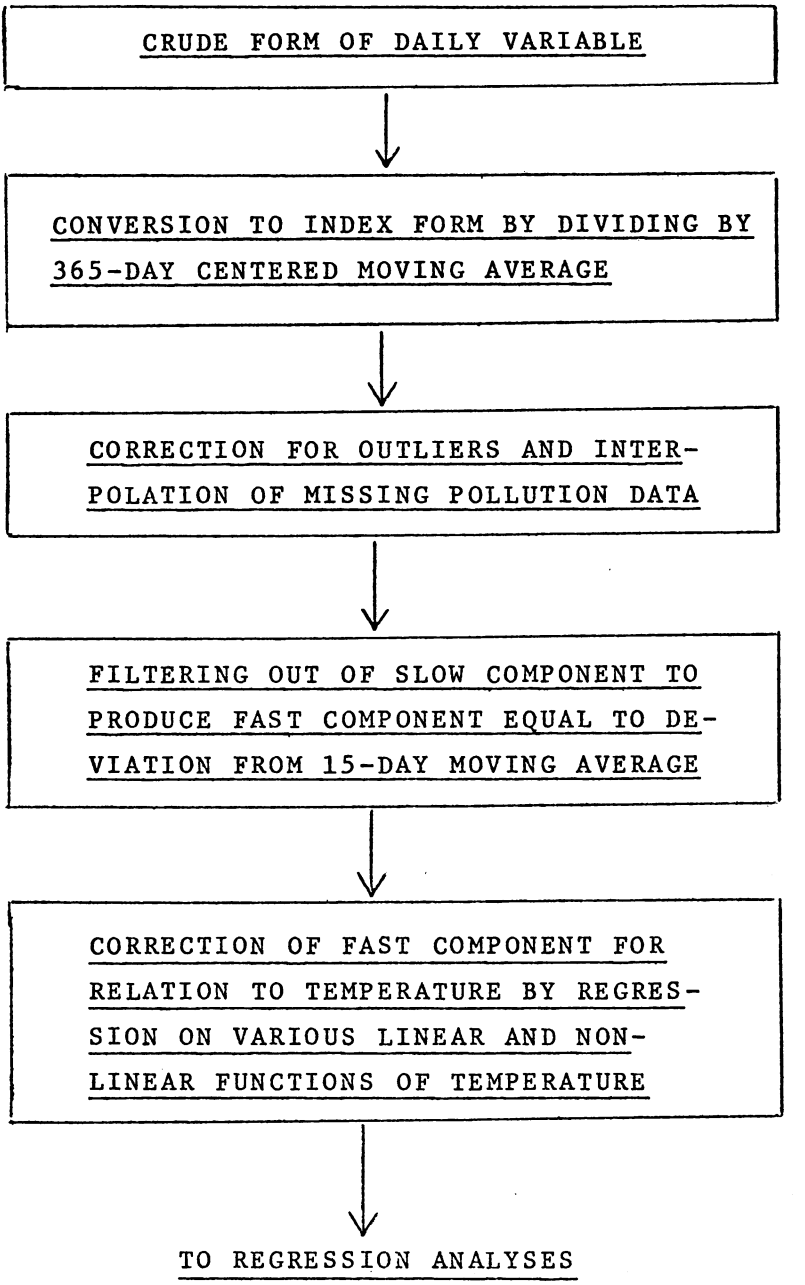

Fig. 1. Adjustment of mortality and pollution variables prior to regression analyses.

TABLE V. COMPARISON OF CORRELATIONS OF SLOW AND FAST COMPONENTS*

	T	P_1	P_2	M_1
T		−0.60	−0.79	−0.67
P_1	*0.34*		0.69	0.52
P_2	*0.34*	*0.55*		0.61
M_1	*0.25*	*0.12*	*0.14*	

*Correlation coefficients between moving averages of crude T and index P_1, P_2, and M_1 are to the right and above the diagonal.

Coefficients between filtered T and filtered index P_1, P_2, and M_1 are in lower left.

M1-Total Mortality and M10-Adjusted Total Mortality have been completed and validated. However, in presenting results some comments will be made on M2-Respiratory Disease, M3-Heart Disease, and M5-Cancer where the initial analyses require further checking. See the final section for additional results completed following the meeting.

Table III gives mean daily values for M1 and M10 and their standard deviations for the three periods 1963 to 1969, 1970 to 1976, and 1963 to 1976.* Figure 1 shows the processing operations which have been applied to these variables. The SO_2 and Smoke Shade measurements at the laboratory station which provided data for earlier studies were extended into mid-June 1975, when the series ended. Similar data were obtained for the period 1971 to 1976 operated by New York State at Roosevelt Island and added to the file. The pollution levels at this station were used to extrapolate the laboratory station series, taking into account that levels at Roosevelt Island station were lower. A multiplier based on the ratio of means of the two stations during the common period January 1971 through May 1975 was used. The correlation matrix for measurements at the two stations is shown in Table IV.

CORRELATION BETWEEN PRINCIPAL VARIABLES

One important reason to filter variables is the differential statistical interaction of the slow and fast components of mortality, pollution, and temperature variables as shown, for example, by correlation coefficients. Table V shows that the slow component of temperature correlates entirely differently with M1 and the pollution variables than does its fast compo-

*Annual data and the seasonal pattern of these variables with monthly means and the standard deviations by month with various corrections are available from the author as Tables A3, A4, and A5.

TABLE VI. MEANS OF MORTALITY AND POLLUTION VARIABLES* CLASSIFIED BY QUARTILES BASED ON SO_2; VARIABLES IN FILTERED INDEX FORM WITH AND WITHOUT TEMPERATURE CORRECTION

Variables	*Temp. corr.*	*Period*	*Quartiles* *1st*	*2nd*	*3rd*	*4th*	*S.E.*
M_1*—Total mortality	Yes	63-69	−0.10	−0.00	0.16	−0.06	0.25
M_1—Total mortality	Yes	70-76	0.05	−0.06	0.19	−0.20	0.27
M_1—Total mortality	Yes	63-76	−0.05	0.01	0.21	−0.25	0.19
M_{10}*—Adj. total mortality	Yes	63-69	−0.27	0.00	0.26	0.01	0.29
M_{10}—Adj. total mortality	Yes	70-76	−0.14	−0.30	0.36	0.08	0.30
M_{10}—Adj. total mortality	Yes	63-76	−0.16	−0.11	0.35	−0.08	0.21
P_1*—SO_2	Yes	63-76	−0.44	−0.13	0.07	0.50	
P_2*—Smoke Shade	Yes	63-76	−0.14	−0.04	0.04	0.14	
M_1—Total mortality	No	63-69	−1.41	−0.88	0.61	1.59	0.30
M_1—Total mortality	No	70-76	−1.34	−0.15	0.06	1.43	0.29
M_1—Total mortality	No	63-76	−1.37	−0.39	0.36	1.41	0.21
M_{10}—Adj. total mortality	No	63-69	−1.36	−1.15	0.90	1.61	0.33
M_{10}—Adj. total mortality	No	70-76	−1.42	−0.27	0.26	1.43	0.32
M_{10}—Adj. total mortality	No	63-76	−1.38	−0.65	0.48	1.57	0.23
P_1—SO_2	No	63-76	−0.49	−0.16	0.03	0.60	
P_2—Smoke Shade	No	63-76	−0.21	−0.06	0.05	0.22	

*Means are referenced to grand means of 100.00 for mortality and 1.00 for pollution. S.E. value is approximate since S.E. values for quartiles vary up to 10% from figure used. See text.

nent with their respective fast components. The coefficients in Table V are abstracted from a more complete matrix.* These tables point to the significant fact that analyzing crude variables by two-month periods is not a suitable substitute for filtering. Thus, deviations P_1 and P_2 correlate with each other with coefficients in the range 0.4 to 0.6 by annual periods and by two-month periods. Crude P_1 and P_2 correlated in this same range for annual periods, but the coefficients drop into the range 0.02 to 0.45 for correlation by two-month periods.

Same-Day Effects

SO_2. No significant statistical association between daily SO_2 and daily mortality is shown in separate and in joint regressions when the variables are corrected more extensively for temperature relations. This was shown for Total Mortality-M1 in Table I, lines 10, 11, and 12 where the estimates for the Mean Standardized Regression Coefficients in a joint regression were −0.23, −0.43, and −0.35 for the periods 1963 to 1969, 1970 to 1976, and 1963 to 1976, respectively. Standard errors for these estimates are 0.43, 0.33, and 0.26, respectively. Results for same-day regressions on SO_2 alone are contained in the more extensive Table IX. Regressions for M10-Adjusted Total Mortality show similar, not statistically significant, associations (see Table X).

Particulates. Association between daily particulate levels as measured by Smoke Shade and daily mortality is positive and statistically significant in regressions of M1 and M10, using filtered, temperature-corrected variables. Details are contained in Tables IX and X. Estimates are lower than those obtained in previous studies, but relative precision of estimates is greater because of the enlarged data base. Estimates are higher in the joint regressions for both SO_2 and Smoke Shade because the SO_2 coefficients are negative. In the regressions of M1 and M10 alone for the full 14-year period, coefficients are 0.71% (S.E. = 0.36) and 1.26% (S.E. = 0.41), respectively. The latter M10 is a more homogeneous category and excludes violence, accidents, etc. as well as many infrequent causes of death.

For both M1 and M10, coefficients are higher in the earlier seven-year period. Differences are not statistically significant, although they correspond to the reduction in Smoke Shade levels between the two periods.

*These data are available from the author as Tables A-6 and A-7.

TABLE VII. MEANS OF MORTALITY AND POLLUTION VARIABLES* CLASSIFIED BY QUARTILES BASED ON SMOKE SHADE*; VARIABLES IN FILTERED INDEX FORM WITH AND WITHOUT TEMP. CORRECTION

Variables	*Temp. corr.*	*Period*	*Quartiles* *1st*	*2nd*	*3rd*	*4th*	*S.E.**
M_1*—Total mortality	Yes	63-69	−0.44	0.19	−0.13	0.38	0.26
M_1—Total mortality	Yes	70-76	−0.69	0.44	0.31	−0.06	0.27
M_1—Total mortality	Yes	63-76	−0.59	0.30	0.10	0.19	0.18
M_{10}*—Total mortality	Yes	63-69	−0.61	0.24	−0.10	0.47	0.29
M_{10}—Total mortality	Yes	70-76	−0.91	0.30	0.50	0.11	0.30
M_{10}—Total mortality	Yes	63-76	−0.74	0.23	0.20	0.31	0.21
P_1*—SO_2	Yes	63-76	−0.22	−0.05	0.05	0.21	
P_2*—Smoke shade	Yes	63-76	−0.29	−0.09	0.05	0.33	
M_1—Total mortality	No	63-69	−2.00	−0.23	0.39	1.82	0.29
M_1—Total mortality	No	70-76	−1.25	−0.45	0.51	1.19	0.29
M_1—Total mortality	No	63-76	−1.58	−0.38	0.50	1.47	0.21
M_{10}—Adj. total mortality	No	63-69	−2.09	−0.33	0.54	1.89	0.32
M_{10}—Adj. total mortality	No	70-76	−1.52	−0.57	0.71	1.39	0.32
M_{10}—Adj. total mortality	No	63-76	−1.80	−0.44	0.63	1.62	0.23
P_1—SO_2	No	63-76	−0.30	−0.9	0.06	0.33	
P_2—Smoke shade	No	63-76	−0.35	−0.11	0.07	0.39	

*Means are references to grand means of 100.00 for mortality and 1.00 for pollution. S.E. value is approximate since S.E. values for quartiles vary up to 10% from figure used. See text.

Alternative Analysis of Same-Day Effects

An alternative relatively simple analysis classified sample days on the basis of the quartile distribution of pollution variables. Means and standard deviations of the mortality variables in each class were then computed. Table VI summarizes such an analysis for which quartiles were computed on the basis of SO_2. Two separate analyses are contained in this table. In one, the filtered temperature-corrected variable $P_1{}^{1*}$ is used as a basis for the quartile distribution and means of the corresponding $M_1{}^{1*}$ and $M_{10}{}^{1*}$ filtered and temperature-corrected variables are shown. In the second, the analysis is the same except that SO_2 is filtered only in determining the quartiles, and the corresponding mortality variables are also filtered but without temperature correction. The analysis is by early, late, and combined periods. For brevity, the quartile mean values are given only for the combined period for SO_2 because those of the two separate periods are approximated well by the combined period. Similar analyses based on Smoke Shade are given in Table VII.

These analyses provide additional insight into regression results and are best examined visually in the plots contained in Figure 2. Both mortality variables show clear positive association with each pollution variable when these are not corrected for temperature, i.e., as pollution increases mortality increases. It is readily seen from the plots of twice the standard errors of the means that this behavior is statistically significant.

On the other hand, plots for temperature-corrected variables show why no statistically significant regression coefficient was found when mortality was regressed on SO_2 alone and also explain why lower estimates were obtained in the second seven-year period when compared with the first. The mean values for the combined 1963 to 1976 period by quartiles suggest that a threshold rather than a linear model would be more appropriate.

The above analyses correspond to regressions on SO_2 and Smoke Shade separately. Analyses were also carried out, using a two-way classification based on the quartiles of the two pollution variables. Table VIII summarizes one of these analyses, namely for $M_{10}{}^{1*}$-Adjusted Total Mortality for the period 1963 to 1976, using filtered, temperature-corrected variables. The marginal values correspond to those which appear toward the bottom of Tables VII and VIII. Inspection of the number of sample days, N, appearing in each of the cells shows the positive correlation between

TABLE VIII. MEANS AND STANDARD ERRORS—TWO WAY CLASSIFICATION OF SAMPLE DAYS BY POLLUTION QUARTILES (See Text)

MARGIN	M −0.75 (0.21) P_1 −0.21 P_2 −0.30 N 1281	0.23 (0.21) −0.05 −0.09 1271	0.20 (0.21) 0.05 0.05 1283	0.31 (0.21) 0.21 0.33 1279	0.00 (0.10) 0.00 0.00 5114
P_1	M −0.95 (0.70) P_1 0.49 P_2 −0.27 N 119	−0.23 (0.52) 0.47 −0.08 218	0.14 (0.38) 0.44 0.06 359	0.03 (0.30) 0.56 0.36 585	−0.08 (0.21) 0.50 0.14 1281
	M −0.63 (0.54) P_1 0.06 P_2 −0.27 N 207	0.49 (0.40) 0.06 −0.09 324	0.56 (0.39) 0.07 0.06 394	0.56 (0.40) 0.07 0.31 350	0.35 (0.21) 0.06 0.04 1275
	M −1.11 (0.42) P_1 −0.14 P_2 −0.28 N 334	0.09 (0.36) −0.13 −0.09 406	0.32 (0.43) −0.12 0.05 321	0.39 (0.53) −0.12 0.30 216	−0.11 (0.21) −0.13 −0.04 1277
	M −0.55 (0.29) P_1 −0.48 P_2 −0.33 N 621	0.45 (0.40) −0.40 −0.09 323	−0.56 (0.55) −0.39 0.05 209	0.83 (0.66) −0.41 0.33 128	−0.16 (0.31) −0.44 −0.14 1281
			P_2		MARGIN

M_{10}* = Adjusted total mortality 1963-1976 temperature corrected.

SO_2 and Smoke Shade, with diagonal elements of the 16-cell array having the largest N in their respective column or row. Careful inspection showed no clear interaction between pollution variables, but provided the detail explaining the marginal totals.

INDEX BEHAVIOR OF SO_2

At this point it is appropriate to comment on the index behavior of SO_2. SO_2 level is widely used as an indicator of ambient air pollution in statistical studies. The reduction of SO_2 levels provides an opportunity to examine the sense in which it serves as an indicator. The previous study showed that the correlation coefficients between crude mean daily SO_2 and Smoke Shade were 0.60, 0.67, and 0.62 for the periods 1963 to 1966, 1967 to 1969, and 1970 to 1972 despite sharp reductions in SO_2 levels in the last period. Corresponding coefficients for filtered variables were 0.59,

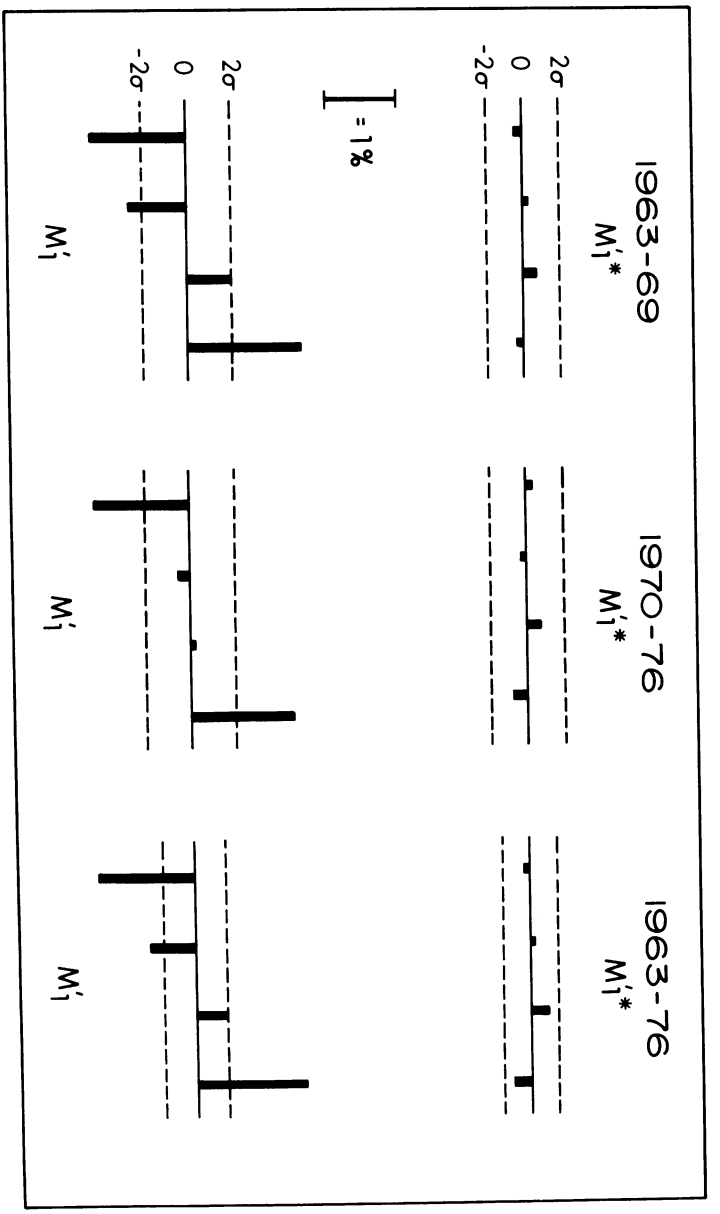

Fig. 2a. Means and standard deviations of M_1 = Total Mortality: Quartile classification based on SO_2.

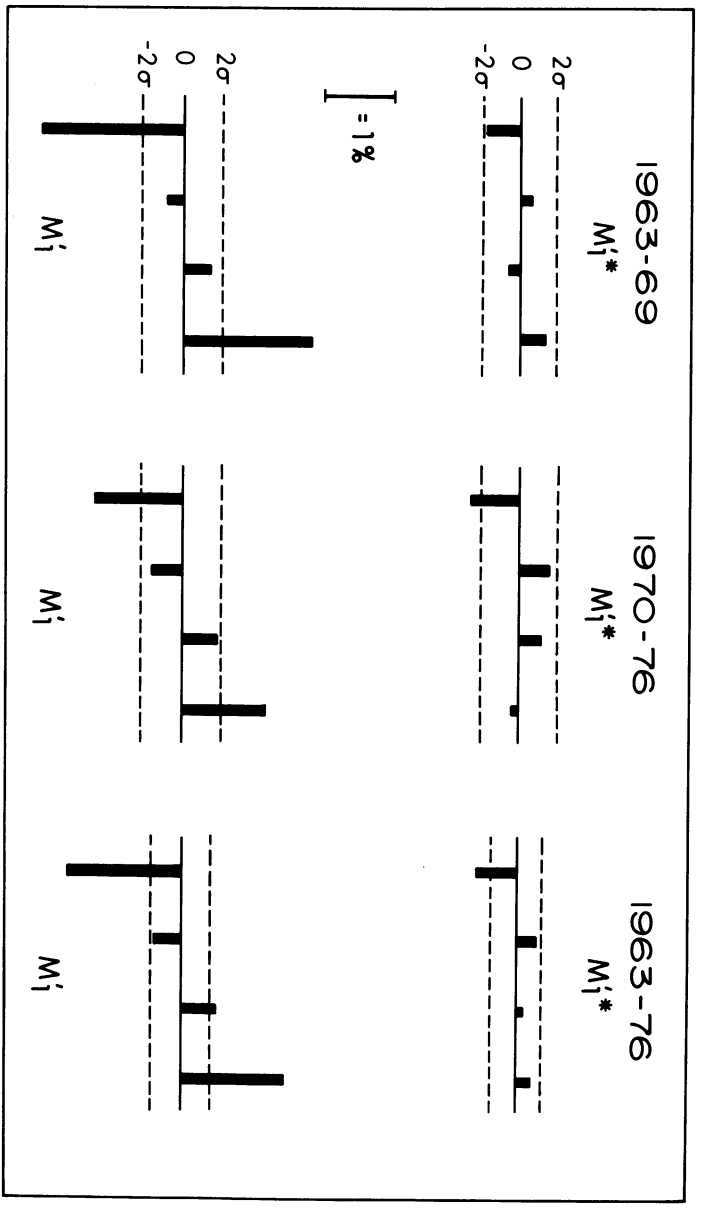

Fig. 2b. M_1= Total Mortality:Quartile classification based on Smoke Shade.

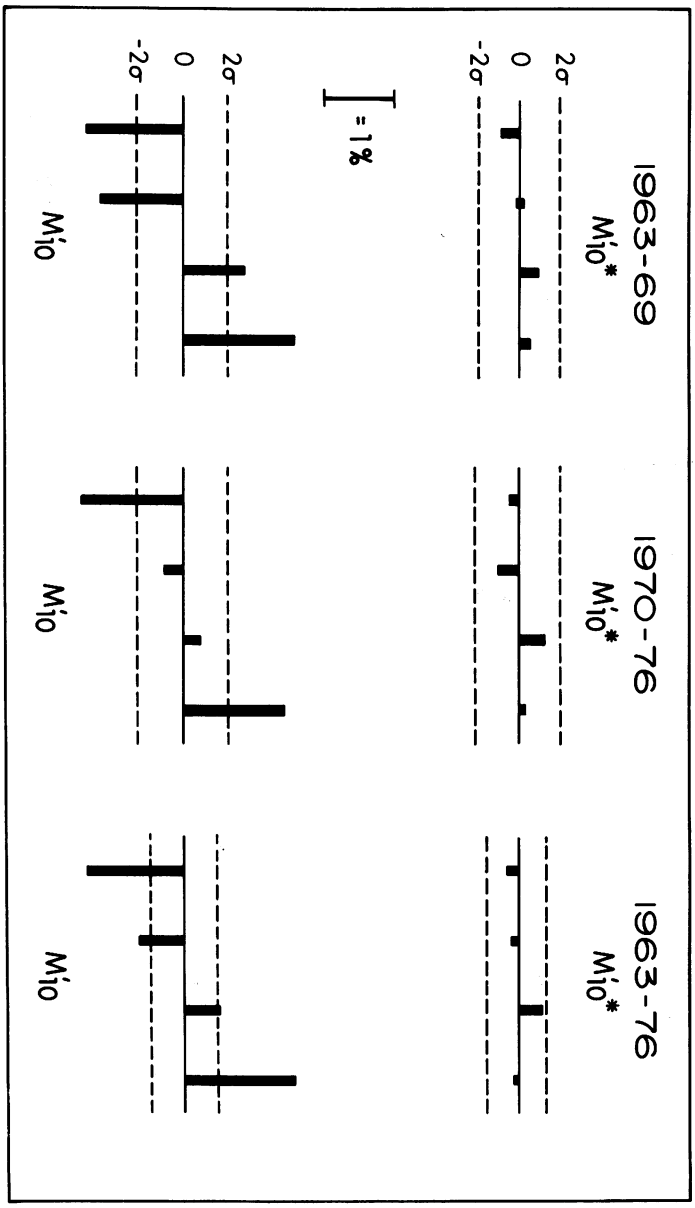

Fig. 2c. M_{10}= Adjusted Total Mortality: Quartile classification based on SO_2.

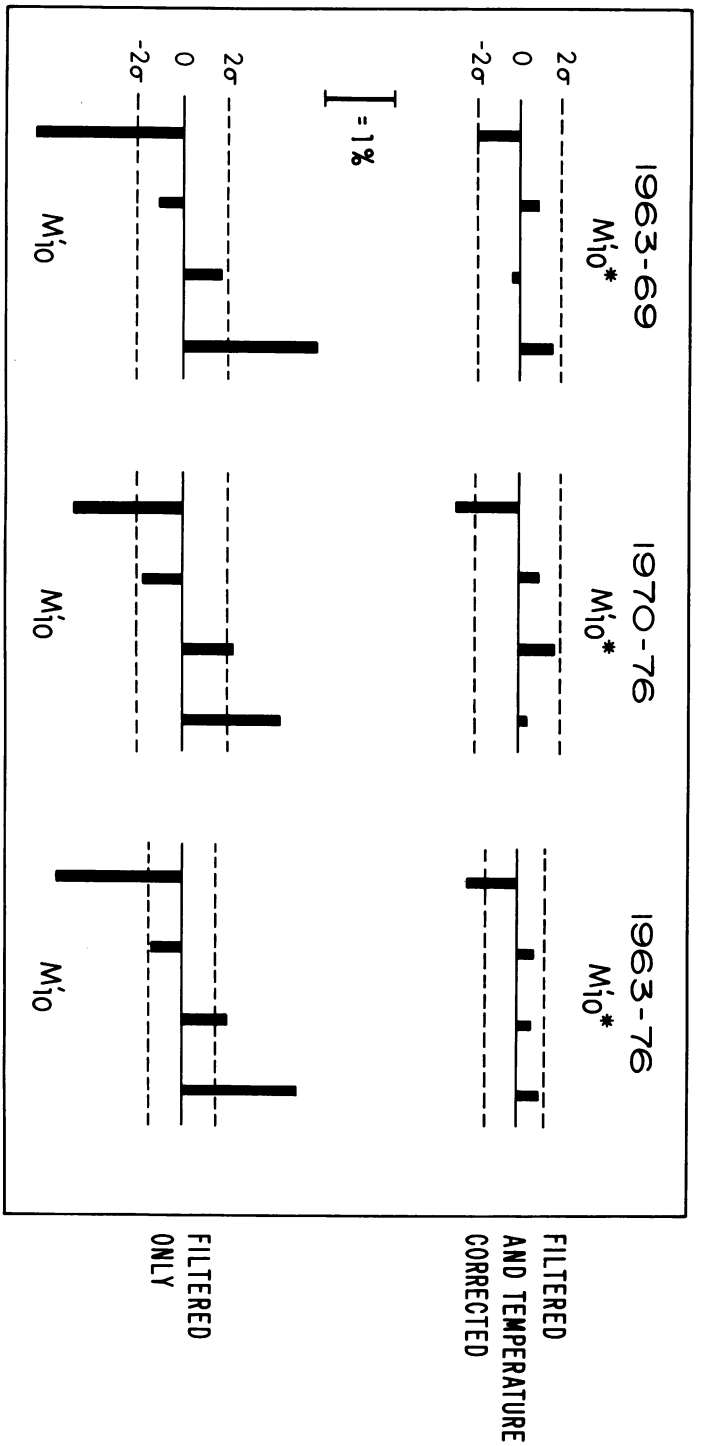

Fig. 2d. M_{10}= Adjusted Total Mortality: Quartile classification based on Smoke Shade.

−2 TO 2	1963-69	0.75(0.87)	2.57(1.20)	0.01(0.96)	2.54(1.32)	2.55(1.26)
−2 TO 2	1970-76	−0.99(0.66)	−0.91(1.13)	−0.88(0.70)	−0.47(1.20)	−1.15(1.18)
−2 TO 2	1963-76	−0.37(0.52)	0.68(0.82)	−0.56(0.56)	0.96(0.88)	0.40(0.86)
0 TO 4	1963-69	−0.06(0.87)	2.34(1.20)	−0.93(0.96)	2.83(1.32)	1.89(1.26)
0 TO 4	1970-76	−0.32(0.66)	1.60(1.13)	−0.71(0.70)	2.02(1.20)	1.31(1.18)
0 TO 4	1963-76	−0.21(0.52)	1.94(0.82)	−0.76(0.56)	2.35(0.88)	1.59(0.86)
−4 TO 4	1963-69	1.52(1.53)	4.19(2.20)	0.57(1.66)	3.56(2.41)	4.13(2.32)
−4 TO 4	1970-76	−0.99(1.22)	−0.33(1.97)	−1.04(1.29)	0.25(2.10)	−0.79(2.06)
−4 TO 4	1963-76	−0.08(0.95)	1.66(1.47)	−0.47(1.01)	1.86(1.57)	1.38(1.54)
0 TO 7	1963-69	−0.67(1.36)	3.12(1.92)	−1.82(1.49)	4.17(2.10)	2.36(2.03)
0 TO 7	1970-76	1.22(1.07)	3.63(1.75)	0.48(1.14)	3.46(1.85)	3.94(1.82)
0 TO 7	1963-76	0.56(0.84)	3.40(1.29)	−0.32(0.90)	3.67(1.38)	3.35(1.35)
−7 TO 7	1963-69	−0.33(2.70)	4.51(4.14)	−1.61(2.86)	5.06(4.38)	3.44(4.39)
−7 TO 7	1970-76	2.15(2.36)	6.78(3.77)	0.90(2.47)	6.70(3.93)	7.60(4.01)
−7 TO 7	1963-76	1.09(1.77)	5.43(2.78)	−0.13(1.86)	5.65(2.91)	5.52(2.95)
0 TO 13	1963-69	−1.23(2.47)	−0.55(3.76)	−1.06(2.64)	0.01(4.04)	−1.05(3.97)
0 TO 13	1970-76	1.42(2.13)	3.49(3.43)	0.74(2.23)	3.29(3.59)	4.03(3.63)
0 TO 13	1963-76	0.37(1.61)	1.88(2.53)	−0.01(1.70)	1.93(2.67)	1.93(2.67)

TABLE IX. DEPENDENT VARIABLE = TOTAL MORTALITY (M1): TEMPERATURE-CORRECTED HIGH-FREQUENCY COMPONENT. SUMMED REGRESSION COEFFICIENTS, IN PERCENT, AND THEIR STANDARD ERRORS (IN PARENTHESES) FOR SEPARATE AND JOINT REGRESSIONS ON TEMPERATURE-CORRECTED HIGH-FREQUENCY COMPONENTS OF SO_2 AND SMOKE SHADE POLLUTION VARIABLES

		Separate		*Joint*		
Lag days	*Period*	*SO_2*	*Smoke*	*SO_2*	*Smoke*	*Combined*
0	1963-69	0.20(0.38)	1.07(0.53)	−0.23(0.43)	1.23(0.61)	1.00(0.55)
0	1970-76	−0.27(0.31)	0.43(0.49)	−0.43(0.33)	0.69(0.53)	0.26(0.51)
0	1963-76	−0.09(0.24)	0.71(0.36)	−0.35(0.26)	0.94(0.40)	0.59(0.37)
0 TO 1	1963-69	0.24(0.50)	1.70(0.70)	−0.45(0.57)	2.09(0.79)	1.63(0.72)
0 TO 1	1970-76	−0.37(0.39)	0.93(0.65)	−0.63(0.42)	1.28(0.69)	0.65(0.67)
0 TO 1	1963-76	−0.15(0.31)	1.31(0.47)	−0.56(0.33)	1.65(0.51)	1.09(0.49)
−1 TO 1	1963-69	−0.17(0.62)	1.49(0.86)	−0.78(0.69)	1.98(0.96)	1.19(0.90)
−1 TO 1	1970-76	−0.82(0.48)	−0.10(0.81)	−0.86(0.51)	0.36(0.86)	−0.49(0.84)
−1 TO 1	1963-76	−0.59(0.38)	0.62(0.59)	−0.82(0.41)	1.10(0.63)	0.27(0.81)
0 TO 2	1963-69	0.24(0.62)	1.95(0.86)	−0.50(0.69)	2.25(0.96)	1.75(0.90)
0 TO 2	1970-76	−0.52(0.48)	0.66(0.81)	−0.71(0.51)	1.03(0.86)	0.32(0.84)
0 TO 2	1963-76	−0.25(0.38)	1.24(0.59)	−0.62(0.41)	1.59(0.63)	0.97(0.61)

−2 To 2	1963-69	1.41(0.97)	4.48(1.34)	0.12(1.07)	4.38(1.47)	−4.50(1.41)
−2 To 2	1970-76	−0.88(0.75)	−1.66(1.28)	−0.59(0.79)	−1.39(1.35)	−1.97(1.33)
−2 To 2	1963-76	−0.07(0.59)	1.14(0.92)	−0.32(0.63)	1.28(0.99)	0.96(0.97)
0 To 4	1963-69	0.29(0.97)	3.22(1.34)	−0.83(1.07)	3.63(1.47)	2.80(1.41)
0 To 4	1970-76	−0.23(0.75)	0.75(1.28)	−0.49(0.79)	1.05(1.35)	0.56(1.33)
0 To 4	1963-76	−0.04(0.59)	1.89(0.92)	−0.57(0.63)	2.18(0.99)	1.61(0.97)
−4 To 4	1963-69	2.75(1.71)	6.99(2.46)	1.18(1.86)	5.98(2.69)	7.16(2.60)
−4 To 4	1970-76	−1.98(1.38)	−2.30(2.23)	−1.72(1.46)	−1.39(2.37)	−3.19(2.38)
−4 To 4	1963-76	−0.23(1.07)	1.74(1.65)	−0.64(1.14)	2.04(1.77)	1.40(1.73)
0 To 7	1963-69	−0.05(1.53)	2.03(2.14)	−0.61(1.66)	2.35(2.35)	1.73(2.27)
0 To 7	1970-76	1.63(1.21)	2.43(1.97)	1.15(1.29)	1.85(2.10)	3.00(2.06)
0 To 7	1963-76	1.02(0.94)	2.29(1.45)	0.49(1.01)	2.03(1.56)	2.53(1.52)
−7 To 7	1963-69	1.40(3.02)	4.34(4.63)	0.66(3.19)	3.83(4.89)	4.43(4.91)
−7 To 7	1970-76	1.48(2.67)	5.75(4.26)	0.45(2.79)	5.88(4.44)	6.33(4.53)
−7 To 7	1963-76	1.36(2.00)	4.54(3.13)	0.49(2.09)	4.45(3.28)	4.94(3.32)
0 To 13	1963-69	−1.85(2.76)	−1.24(4.21)	−1.67(2.95)	−0.64(4.51)	−2.32(4.44)
0 To 13	1970-76	1.68(2.41)	2.58(3.89)	1.05(2.53)	2.20(4.06)	3.25(4.11)
0 To 13	1963-76	0.26(1.81)	1.21(2.84)	−0.06(1.91)	1.23(3.00)	1.17(3.63)

TABLE X. DEPENDENT VARIABLE = SUM OF M2 THRU M5 (M10); TEMPERATURE-CORRECTED HIGH-FREQUENCY COMPONENT.SUMMED REGRESSION COEFFICIENTS, IN PERCENT, AND THEIR STANDARD ERRORS (IN PARENTHESES) FOR SEPARATE AND JOINT REGRESSIONS ON TEMPERATURE-CORRECTED HIGH-FREQUENCY COMPONENTS OF SO_2 AND SMOKE SHADE POLLUTION VARIABLES

		Separate		*Joint*		
Lag days	*Period*	*SO_2*	*Smoke*	*SO_2*	*Smoke*	*Combined*
0	1963-69	0.35(0.42)	1.53(0.60)	−0.24(0.48)	1.69(0.68)	1.45(0.61)
0	1970-76	0.13(0.35)	1.04(0.56)	−0.13(0.38)	1.12(0.60)	0.99(0.58)
0	1963-76	0.21(0.27)	1.26(0.41)	−0.17(0.29)	1.36(0.45)	1.20(0.42)
0 To 1	1963-69	0.44(0.56)	2.52(0.70)	−0.52(0.63)	2.87(0.88)	2.34(0.81)
0 To 1	1970-76	−0.21(0.44)	1.16(0.73)	−0.51(0.47)	1.45(0.78)	0.94(0.76)
0 To 1	1963-76	0.03(0.35)	1.77(0.53)	−0.50(0.37)	2.08(0.58)	1.58(0.55)
−1 To 1	1963-69	0.20(0.69)	2.52(0.96)	−0.74(0.77)	2.99(1.08)	2.25(1.01)
1 To 1	1970-76	−0.72(0.54)	−0.18(0.91)	−0.70(0.57)	0.23(0.97)	−0.48(0.95)
−1 To 1	1963-76	−0.38(0.42)	1.04(0.66)	−0.69(0.46)	1.46(0.71)	0.77(0.69)
0 To 2	1963-69	0.60(0.69)	3.00(0.96)	−0.47(0.77)	3.28(1.08)	2.82(1.01)
0 To 2	1970-76	−0.39(0.54)	5.36(0.91)	−0.51(0.57)	2.62(0.97)	0.11(0.95)
0 To 2	1963-76	−0.04(0.42)	1.56(0.66)	−0.48(0.46)	1.82(0.71)	1.34(0.69)

0.65, and 0.55. In this sense it serves as an indicator. Even more dramatic index behavior is demonstrated in Figures 2A-2D, where "effects" estimated by using filtered mortality and SO_2 variables without temperature correction are found to be approximately the same in the period 1963 to 1969 as in the period 1970 to 1976 although levels are on the average four to five times higher in the earlier period.

In the previous study, regressions on SO_2 seemed to index Smoke Shade just as they do in Figures *2a-2d* for data without temperature correction. With the introduction of more refined corrections for temperature, significant effects are no longer associated with SO_2 when it is the sole measure of pollution.

Regressions with Lags

Tables IX and X summarize the results obtained using filtered temperature-corrected variables in separate and joint regressions of the mortality variables on the two pollution variables. The regressions include causal and noncausal models, with 10 particular lag models providing 10 coefficients for each independent variable in each regression. Separate regressions for the early, late, and combined periods were carried out. In the joint regressions for M1, only two of 10 SO_2 coefficients were positive in the early period and only three out of 10 in the late period. For the combined period, none of the 10 computed coefficients were positive. In the separate regressions of M1 on SO_2 alone, 11 out of 30 coefficients were positive. Only one of these had a t value greater than 1 ($t = 1.14$). In contrast, three of the negative values had absolute values of t greater than 1, the largest being 1.67.

In contrast, in the joint regressions of M1 on the two pollutant variables, 29 out of 30 Smoke Shade coefficients were positive, and in the separate regressions 26 out of 30 were positive. For many of these coefficients $t > 1.94$ with $a > 0.025$. The pattern for M10 regressions was similar, showing somewhat greater positivity for SO_2. However, the highest t value for a causal model was 1.35. Unless one is prepared to accept the hypothesis that the presence of SO_2 tends to reduce mortality, attention should be restricted to regressions on Smoke Shade alone.

For such regressions for both M1 and M10, the highest coefficients appear in the noncausal models. These are the models in which the pollution variables subsequent to mortality are included in the analysis as well as the previous pollution.

TABLE XI. ESTIMATES OF PERCENT PREMATURE DEATHS AND STANDARD ERRORS FOR SO_2 AND COHs FROM REGRESSIONS OF ALL-COUNTY MORTALITY ON POLLUTION AND DAY OF WEEK BY SINGLE STATION AND MULTISTATION MODEL FILTERED, FULL TEMPERATURE CORRECTED VARIATES. TOTAL MORTALITY 1972 TO 1975.

	Pollutant variate in model			
	SO_2 alone	*COHs alone*	*SO_2 and COHs together*	
Station	*% due to SO_2*	*% due to COHs*	*% due to SO_2*	*% due to COHs*
Hazelwood	−1.66	−0.23	−2.46	1.38
	(0.81)†	(0.86)	(1.03)†	(1.09)
Bellevue	−0.54	−0.98	−0.23	−0.88
	(0.96)	(1.05)	(1.04)	(1.14)
Logans Ferry	0.56	−0.12	0.57	−0.20
	(0.45)	(0.87)	(0.45)	(0.87)
Multistation model	−1.27	−0.72	−2.35	1.03
	(1.25)	(1.15)	(1.48)	(1.39)

*Significant, $P < .10$
†Significant, $P < .05$
Preliminary results of Pittsburgh study furnished in personal communication to Herbert Schimmel from S. Mazumdar and N. Sussman.

Tables IX and X contain only the sums of the coefficients for the individual days. To gain further understanding of these results the individual coefficients were examined. Figures 3*a* and 3*b* plot coefficients by individual days with causal models on the left side and noncausal or symmetrical models on the right. The figures show that as additional days are added, estimates for the earlier days change very little.

It seems that the results shown in the figures are not consistent with a definite lag effect, that great care should be exercised in using the results contained in Tables IX and X, and that at most results based on same-day and previous-day pollution should be considered. If this is done, somewhat higher estimates of effects associated with Smoke Shade are obtained, namely 2.52% (S.E. = 0.78) for 1963 to 1969 and 1.16 (S.E. = 0.73) for 1970 to 1976. While the estimates for the early period are more than twice as large as for the later period, the difference is not statistically significant.

PRELIMINARY RESULTS FROM PITTSBURGH STUDY

Dr. Sati Mazumdar and associates of the University of Pittsburgh have applied a similar methodology to data for that city.[5] Their data cover four

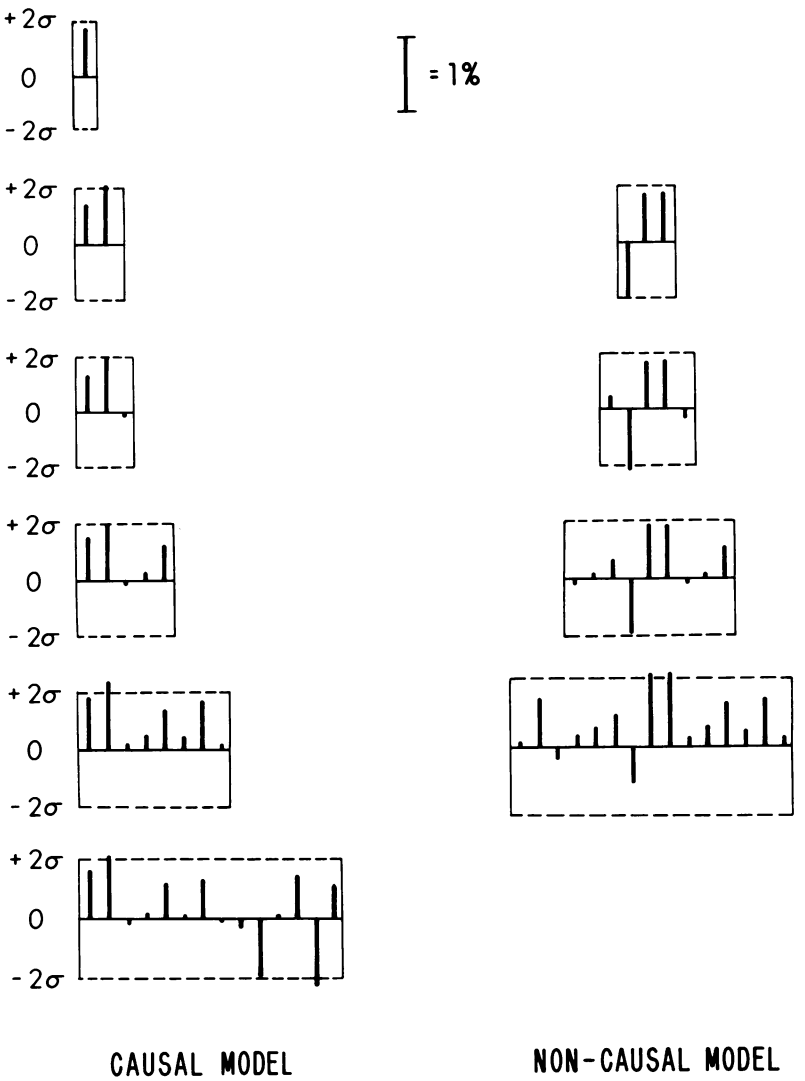

Fig. 3a. Response of total mortality—M_1 to Smoke Shade P_2 variables filtered and temperature corrected.

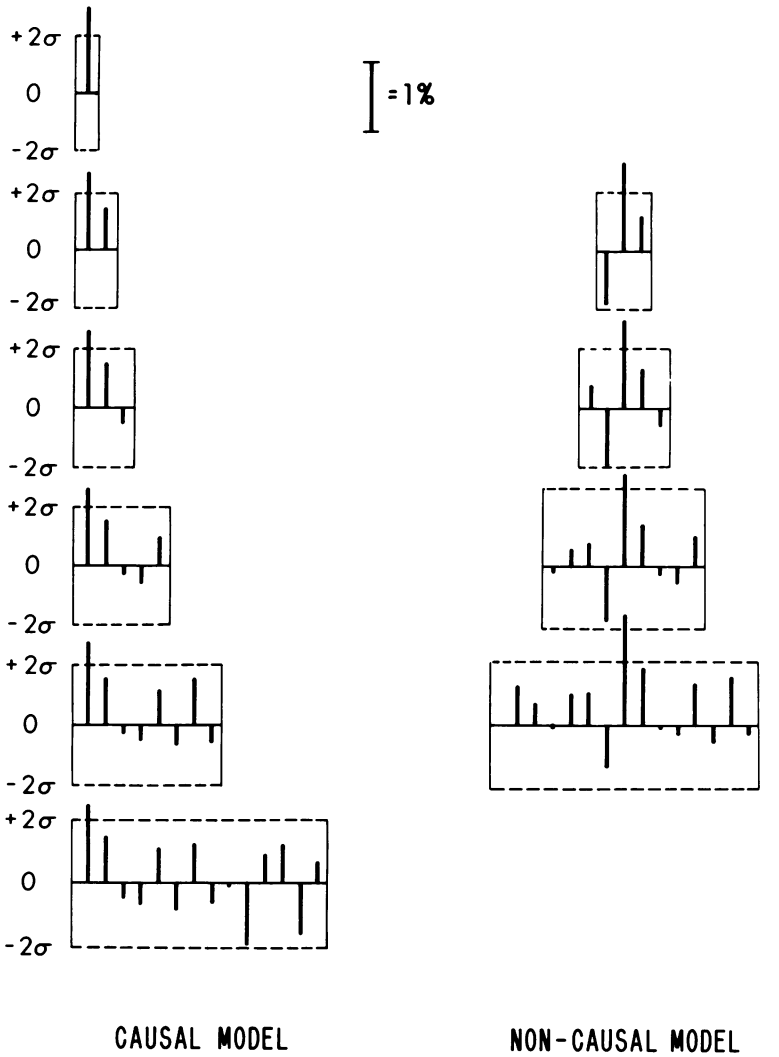

Fig. 3b. Response of adjusted total mortality—M_{10} to Smoke Shade P_2 variables filtered and temperature corrected.

TABLE XII. ESTIMATES OF PERCENT PREMATURE ILLNESS AND STANDARD ERRORS FOR SO_2 AND COHs FROM REGRESSIONS OF ALL-COUNTY MORBIDITY ON POLLUTION AND DAY OF WEEK BY SINGLE STATION AND MULTISTATION MODEL FILTERED, FULL TEMPERATURE CORRECTED VARIATES. TOTAL MORBIDITY 1972 TO 1975.

	Pollutant variate in model			
	SO_2 alone	*COHs alone*	*SO_2 and COHs together*	
Station	*% due to SO_2*	*% due to COHs*	*% due to SO_2*	*% due to COHs*
Hazelwood	0.39	0.88	−0.20	1.01
	(0.45)	(0.47)*	(0.57)	(0.60)*
Bellevue	1.70	2.42	1.00	2.00
	(0.53)†	(0.58)†	(0.57)*	(0.62)†
Logans Ferry	−0.03	1.87	−0.19	1.89
	(0.25)	(0.47)†	(0.25)	(0.48)†
Multistation model	1.70	2.67	0.90	2.24
	(0.69)†	(0.63)†	(0.81)	(0.76)†

*Significant, $P < .10$
†Significant, $P < .05$
Preliminary results of Pittsburgh study furnished in personal communication to Herbert Schimmel from S. Mazumdar and N. Sussman.

years only, as compared to the 14 for the New York study, but their study embraces morbidity as well as mortality variables, uses pollution measurements from three stations, and presents only same-day regressions. Preliminary results show poor correlation between SO_2 measurements at different stations and poor correlation between SO_2 and Smoke Shade at the same stations. Correlation between Smoke Shade measurements at different stations is less than that shown by the laboratory station with the New York City monitoring network in the studies by Dr. Inge Goldstein.[9]

Regressions of both mortality and morbidity on SO_2 and Smoke Shade give negative coefficients for SO_2 far larger than those just presented in Tables IX and X. Coefficients for Smoke Shade are positive and statistically significant, as may be seen from Tables XI and XII.

Discussion and Conclusions

EVALUATION OF STATISTICAL RESULTS

The analyses thus far show that as the analysis is refined to eliminate adventitious relations between slow components and to correct for local (in time) weather relations, no statistically significant associations between

mortality and SO_2 are present. The association between particulates, as measured by Smoke Shade, and mortality is small and of moderate statistical significance.

It is not expected that these observations will be seriously altered by the considerable additional analyses which remain to be completed. It will be useful, however, to list these additional analyses and to comment on what they might show. Analyses are to be carried out on the more detailed mortality classifications defined in Table II. Despite loss of precision as the size of the mortality class is reduced, interesting results may emerge along two lines as suggested by preliminary analyses. Statistically significant higher effects may be found associated with the mortality classes 60 + years, and it is likely that the sum of regression coefficients for a causal model for respiratory disease (M_2) will show a systematic increase with lags. How much of this represents the association between particulates and mortality and how much is an analysis artifact due to the known complex response of respiratory disease to weather changes will be difficult to determine.

Threshold models will be tested as outlined in Section II. They may better fit the data. I suspect the principal model which may serve will be of the form $(P_2)^x$, i.e., some power of Smoke Shade. Models where the influence of varying filter-band widths are examined will provide interesting results and should show to what extent the coefficients estimated thus far are stable estimates of day-to-day (in time) relations or residual effects of slower components retained in the filters thus far employed. We note, for example, that the filter employed in the 1963 to 1968 study gave higher same-day estimates than did that in the 1963 to 1972 study, while the form of temperature corrections remained the same.

Finally, analyses will be made, using pollution measurements from seven stations instead of one station, using methods described in Appendix E which have already been applied to Pittsburgh data. According to the considerations set forth in Appendix E, if there is a uniform positive association between pollution and mortality, introduction of the additional pollution variables into the regression equation should give higher estimates of effects. The standard error of estimate will increase if this should occur so as not to make the result very significant statistically (e.g., by an F test). I believe that the single-station problem is a real one, but not particularly in the case of New York City, and was pressed with respect to our previous study for other than statistical reasons.

Importance of Replication

Having observed effects diminish and disappear as greater care was introduced into the analysis, I suspect that the residual effect still observed for particulates may well be an artifact of the New York data base, particularly of weather effects which appear to be far stronger and statistically significant than those associated with pollution. Results of the preliminary analyses of the Pittsburgh mortality, morbidity, and pollution data provide surprising confirmation considering the short period covered and the fragility of such analyses. It would be extremely useful to have similar analyses for cities such as London and Chicago, where weather patterns differ. Considering the prominent and necessary part played by weather corrections in these studies, Mexico City would be of special interest because of its moderate weather changes.

Interpretation and Implications of New Results

In earlier sections it has been emphasized that acute episodes such as those of London in the 1950s and 1960s and the pollution exposure attendant upon cigarette smoking provide evidence of a causal association between pollution at sufficiently high levels and adverse health effects. Broad statistical studies such as the present one and many similar ones should not be given causal interpretation. However, they should be analyzed as though they might be used in a causal context. When effects are weak they must be validated in a larger context of other types of evidence. Interpretation is difficult. The response is ill defined, whether as a mortality or morbidity measure. The agent is a variable which is brought into the analysis representing all possible relevant variables with which it is correlated and have not been included in the analyses.

Additional Statistical Results

Between the time of the meeting and printing of the proceedings of the symposium, a great deal of additional data became available in accordance with the program described in Section II. These included results from regressions, with lags for classes M2 to M7 and details of selected analyses by race, sex, and age. Regression results were obtained for M1 Total Mortality, when white deaths are further classified by sex and age. Similar data were calculated for M10 Adjusted Total Mortality. These

regressions are based on filtered and temperature-corrected pollution variables. The mortality variables are filtered only. As discussed in connection with Table I of the report, this gives estimates of regression coefficients which are almost identical with those obtained, using filtered and temperature-corrected variables. Standard errors of estimate, however, are about 10% higher.

These new results are entirely consistent with the results previously discussed which show no statistically significant associations between daily SO_2 and mortality levels. As to the smoke association, there is also no consistent statistically significant difference in regression coefficients between mortality groups classified by age and sex. In the early period there is a definite suggestion of greater effect in the older groups. This is reversed in the second period.

Acknowledgments

This report includes partial results of a study now in progress under a contract with the Electric Power Research Institute. Dr. Peter Bloomfield of the Department of Statistics, Princeton University, and Dr. T. J. Murawski of the New York State Department of Health served respectively as consultants on statistical and medical questions and participated actively in the development and review of the study. Mr. Clifton Lancaster designed and carried out most of the computer programming and assisted in statistical design and analysis. Dr. Sati Mazumdar of the Department of Statistics, Graduate School of Public Health, University of Pittsburgh, collaborated closely in the development of the multistation model and in testing its usefulness with data developed in a companion study for Pittsburgh. This work is indebted to many others and most particularly Dr. Leonard Greenburg, now retired, an outstanding pioneer student of environmental and industrial health sciences. Acknowledgements would not be complete without mentioning the cooperation of the New York City Departments of Health and Air Resources, which supplied and continue to supply basic data on New York City mortality and air pollution essential to our studies. These acknowledgements notwithstanding, the author has sole responsibility for this report and most particularly its more general observations and conclusions.

Appendix B

Correction for Outliers

A few extreme values for some of the variables may have a disproportionate influence on means, standard deviations, regression coefficients, and the like. To reduce such influence, the pollution and mortality series (not the temperature) were subjected to winsorization. This involves calculating bounds beyond which any values are considered to be outliers; the high outliers are then pushed back to the upper bound and the low to the lower bound. The winsorized series, then, differs from the original series at only a few points.

Index pollution and mortality series were broken up by months (and further by two periods for the laboratory station pollution series). Within each month, means and standard deviations are calculated and the bounds set at the mean ± 3 standard deviations. The numbers and percentages of outliers determined in this way are given below.

Table BI

	Lab. SO_2	*Roos. Is. SO_2*	*Lab. smoke*	*Roos. Is. SO_2*	*M1*	*M10*
Number of outliers	33	18	14	12	38	31
Total length of series	4,547	2,192	4,534	2,192	5,114	5,114

Outliers "caught" ranged from 0.00 to 5.77 for the various pollution series and from 0.76 to 2.47 for the mortality series. The upper bounds applied (depending on variable and month) ranged up to 5.07 for the pollution variables and 1.50 for the mortality variables. Lower bounds ranged down to 0.01 for pollution and 0.55 for mortality. Twenty-seven percent of the pollution outliers and only one of the mortality outliers were low.

Since the mortality series was derived from the actual counting of dated death certificates, extreme values are not by themselves suspect. The pollution series, however, may contain instrumental errors. Of interest in this regard is whether the outliers discovered tend to be preceded and followed by other outliers. A table of runs for the outliers of the four pollution series is given below.

Table BII

Length of run	*Lab. SO_2*	*Roos. Is. SO_2*	*Lab. smoke*	*Roos. Is. smoke*
1	22	10	5	10
2	4	1	3	1
3	1	2	1	0

Interpolation of the Pollution Series

Gaps in the pollution series were filled in before analysis. This was done in the following way. Suppose the series y has a gap in it from time a to time b; in other words, y(a) and y(b) are available but all of y(a + 1),...y(b − 1) are missing. We wish to interpolate the series y over this gap using a segment of another series z(a), ... z(b). The interpolated values y(a + 1), ... y(b− 1) are given by the formula

$$\log y(t) = \log z(t) - \frac{t-b}{b-a}\log y(a) - \frac{a-t}{b-a}\log y(b)$$
$$+ \log z(t) + \frac{t-b}{b-a}\log z(a) + \frac{a-t}{b-a}\log z(b)$$

for t = a + 1, ... b − 1.

Note that the interpolated values are the same if z is multiplied by any (positive) constant.

A summary of the interpolations done follows. Whenever possible, gaps in the SO_2 series for one station were filled, using the SO_2 series for the other station, and likewise for gaps in Smoke Shade.

A day's average was considered missing if it was based on fewer than four readings. A few other averages at the beginning and end of gaps were deleted if they seemed aberrant.

In the case of a few long gaps, the interpolation procedure given above was slightly modified to reduce the influence of the values of the endpoints upon the interpolated segment.

Additional Pollution Data Holdings for Stations 1,3,5,10,11 (New York City)

Smoke Shade: 10/1/70 through 6/30/76. All five stations contain many stretches of missing data.

Sulfur dioxide (peroxide method):

station 1 from 7/1/68
station 3 from 9/3/68
station 5 from 9/3/68
station 10 from 11/29/68
station 11 from 8/1/68

Until about 8/28/75 all five stations contain many stretches of missing data and no Sunday readings were taken. After this time readings were taken on no more than two days a week. Readings end 1/13/77.

Appendix C

Summary of Interpolations Done

This table summarizes the interpolations done for the two Roosevelt Island pollution series from 1971 on and for the two laboratory station pollution series from 1973 on. To the left are the number of gaps for each series by length of gap, and to the right are the sums of gap lengths according to interpolation source.

Table CI

Laboratory station SO_2

Gaps		*Sums*
Length	*Number*	32 total
1	3	31 from Roosevelt Island SO_2
29	1	1 from lab Smoke Shade

Laboratory station Smoke Shade

Gaps		*Sums*
Length	*Number*	12 total
1	6	12 from Roosevelt Island Smoke Shade
2	3	

Roosevelt Island SO_2

Gaps		*Sums*
Length	*Number*	105 total
1	16	70 from interpolated Lab SO_2
2	11	31 from Roosevelt Island Smoke Shade
3	8	3 from station 3 Smoke Shade
4	3	1 from station 1 Smoke Shade
5	1	
10	1	
16	1	

Roosevelt Island Smoke Shade

Gaps		*Sums*
Length	*Number*	61 total
1	12	56 from interpolated laboratory Smoke Shade
2	4	4 from station 3 Smoke Shade
3	3	1 median of day before and day after
4	2	
10	1	
14	1	

Appendix D

Temperature Corrections

To remove the influence of temperature, the mortality and pollution variables were regressed on certain temperature functions and residuals obtained.

The specific temperature functions employed were:

1) Same-day temperature.
2) Previous-day temperature, i.e., 1st lag.
3) Average of 2nd and 3rd lags.
4) Average of 4th to 6th lags.
5) Average of 7th to 13th lags.
6) The square of the positive deviation of same-day temperature from its expected temperature
7) Same as 6 for negative deviations.
8) The square of the positive deviation of the average of same day and previous from their expected value.
9) Same as 8 above for negative deviations.

Temperature used here was defined as mean daily temperature. All the mortality, pollution, and temperature variables were filtered before being entered into the temperature-correction regressions. The initial protocol also included separate variables for the summer based on a straight-line prediction of excess temperature beginning May 1 at 70° F. and ending August 31 at 80° F. This line had been determined empirically as a result of visual inspection of 10 years of data. An examination of July data showed that these correlated at the 0.99 level with some variables from groups six through nine and they were excluded. Also, first-order terms corresponding to six through nine were tested and excluded because they did not add significantly to the R^2, i.e., the proportion of variance accounted for and were highly collinear with 1 and 2.

Since the ultimate regressions of mortality on pollution were to employ filtered index variables, the corrections were carried out with filtered variables, i.e., filtered temperature variables and filtered index mortality and pollution variables. As explained in the text, the temperature-correction regressions were carried out month by month for data pooled over the years of the study.

Table D1 compares the same-day correlation coefficients with the multiple correlation coefficients obtained employing variables 1 through 9. The

TABLE DI—CORRELATION (r) AND MULTIPLE CORRELATION (R) COEFFICIENTS WITH TEMPERATURE BY MONTHS, 1963-1976. (BASED ON FAST COMPONENTS. SEE TEXT.)

	M_1—Total mortality		*P_1—SO_2*		*P_2—Smoke Shade*	
	r	*R*	*r*	*R*	*r*	*R*
January	0.18	0.34	0.28	0.49	0.39	0.55
February	0.12	0.18	0.34	0.53	0.32	0.53
March	0.21	0.25	0.40	0.55	0.39	0.49
April	0.16	0.24	0.42	0.49	0.34	0.34
May	0.32	0.40	0.39	0.47	0.30	0.39
June	0.52	0.56	0.40	0.46	0.33	0.37
July	0.55	0.72	0.46	0.51	0.36	0.38
August	0.50	0.54	0.52	0.56	0.42	0.50
September	0.37	0.42	0.48	0.56	0.40	0.52
October	0.14	0.28	0.27	0.52	0.36	0.52
November	0.15	0.34	0.15	0.50	0.24	0.39
December	0.11	0.32	0.28	0.52	0.39	0.54

corrected mortality and pollution variables, it is seen, might give substantially different results when same-day mortality is regressed on same-day pollution, using only same-day temperature as a correction or control variable.

Finally, the predicted temperature employed for defining variables 6 through 9 was computed by a least square fit of a sine wave to the 14-year data. A comparison of daily averages for New York City for a 106-year period showed they were fitted closely by a sine wave. It was thought best to use a prediction based on the period itself because of long-term changes in the weather.

Table D1 shows that the temperature corrections remove approximately 25% of the variance of the pollution variables, i.e., $R \sim 0.5$, $R^2 \sim 0.25$. The proportion of variance removed from the total mortality variable shows marked seasonal behavior, reaching a high of about 50% in July, gradually declining to about 10% for October through January, remaining less than 10% during the next three months, and then rising during May and June to reach the peak of 50% in July.

An examination of the coefficients of the pollution regressions show that for both SO_2 and Smoke Shade throughout the year, the coefficients are positive for same-day temperatures and negative for lag-day temperatures. In contrast, the coefficients for various mortality variables show more complex patterns.

TABLE DII—SIGNS OF REGRESSION COEFFICIENTS USED FOR TEMPERATURE CORRECTIONS AND THEIR F VALUE FOR DELETION. 1963-1976 DATA (SEE TEXT)

Mo.	*Dependent variable*	*Temperature variable* 1	2	3	4	5	6	7	8	9
1	M_1	+ 2.6	*	−13.2	*	*	*	*	*	*
1	P_1	+35.9	36.2	*	*	*	−2.3	*	*	*
1	P_2	+50.5	−14.6	−2.3	*	*	*	+2.2	*	*
2	M_1	*	*	−2.6	*	*	*	*	+3.1	−2.8
2	P_1	+26.6	−18.3	−10.5	−2.2	−2.1	*	*	*	*
2	P_2	+32.3	−15.9	−17.0	*	*	−2.2	+4.5	*	*
3	M_1	*	*	*	*	*	*	*	*	*
3	P_1	+27.6	−3.5	−2.3	−4.5	*	+4.4	*	7.9	+5.0
3	P_2	+28.7	*	−4.1	−5.5	*	*	*	−5.4	+2.5
4	M_1	+5.7	−3.3	*	*	*	*	*	*	*
4	P_1	+5.6	−10.2	−3.1	*	*	+2.4	*	*	−2.7
4	P_2	*	*	−9.3	*	*	+5.1	*	*	−2.1
5	M_1	*	*	*	*	*	*	*	*	*
5	P_1	43.3	−8.5	−4.2	*	*	*	+8.0	*	*
5	P_2	+16.4	*	−7.8	*	*	*	+3.9	*	*
6	M_1	+8.2	*	*	*	*	*	*	+2.8	*
6	P_1	+21.9	−8.3	*	−2.9	*	*	+2.8	*	−2.3
6	P_2	+17.4	*	*	+3.6	+5.0	*	*	*	*
7	M_1	+12.4	*	*	*	*	−3.6	*	+40.3	*
7	P_1	48.5	−2.8	*	*	*	−2.4	+8.7	*	−3.6
7	P_2	+8.1	*	−2.7	−6.8	−4.9	*	*	*	*
8	M_1	+6.1	*	*	*	*	*	−2.1	*	*
8	P_1	+35.0	−6.5	−5.4	−10.9	−4.2	*	+3.8	+2.3	*
8	P_2	+9.7	10.8	−3.3	−3.5	*	*	*	*	*
9	M_1	+21.6	*	−3.1	*	*	*	+2.9	*	*
9	P_1	+25.9	*	−3.8	*	*	+2.5	*	*	+2.0
9	P_2	+37.8	−9.2	−8.5	*	*	−8.6	*	+9.9	*
10	M_1	*	*	−10.5	*	*	*	*	*	*
10	P_1	+8.4	−25.0	−5.5	−3.5	−9.5	*	*	*	*
10	P_2	+16.7	−22.6	−14.2	*	*	*	*	+4.7	*
11	M_1	−2.3	*	−10.8	*	*	*	*	*	*
11	P_1	+3.1	−19.7	−14.9	−8.0	−6.2	*	*	*	*
11	P_2	+6.3	−8.1	−8.4	−3.1	−2.1	*	*	*	*
12	M_1	*	*	−13.1	−4.8	*	*	*	*	*
12	P_1	+16.2	−14.3	−3.5	*	*	*	*	*	+2.1
12	P_2	+28.4	+18.2	+6.9	*	*	*	*	*	*

*Denotes coefficient where $F < 2.0$

Appendix E

Note on Estimates of Effects Based on Pollution Measurements from One or Several Stations

Hypothetical model. To illustrate the problems of utilizing pollution measurements, a simplified idealized model will be assumed. Only a single type of pollutant (denoted by P) will be present. The area under study is assumed to be divided into K districts, each with a daily mortality variable M_k and a characteristic daily pollution level P_k. It is assumed that the pollutant has an additive effect on mortality ΔM_k proportional to the mean mortality and the amount of pollutant present. Let the relation be described by the following:

$$\text{Equation E1.} \qquad \Delta M_k = a_k P_k \overline{M_k}$$

Note that unlike the pollution variables employed in the body of this report, which were standardized so that the mean value is equal to one, the P_k variables here are in absolute units to take into account the fact that mean values may vary from district to district. The average daily effect for the period of study will be given by

$$\text{Equation E2.} \qquad \Delta M_k = a_k P \overline{M_k} \ .$$

If M represents the total mortality for the area so that

$$\text{Equation E3.} \qquad M = \sum M_k; \qquad \Delta M = \sum \Delta M_k$$

and if W_k is the proportion of total mortality found in district K then

$$\text{Equation E4.} \qquad \begin{aligned} \Delta M &= \sum a_k \overline{M_k} P_k \\ &= \overline{M} \sum a_k W_k P_k \end{aligned}$$

where $W_k = M_k/M$. Also the average proportional effect may be written

$$\text{Equation E5.} \qquad \frac{\overline{\Delta M}}{\overline{M}} = \sum a_k W_k \overline{P_k}$$

Further, if it is desired to express this in terms of a mean weighted coefficient a and mean pollution P, then

$$\frac{\overline{\Delta M}}{\overline{M}} = \bar{a}\bar{P}$$

Equations E6.

$$\bar{P} = \Delta W_k \bar{P}_k$$

$$\bar{a} = \Delta\, a_k W_k \bar{P}_k / \bar{P}$$

If the a_k are all equal then they are equal to $\bar{a}$.

Estimating the coefficients a_k. If the M_k and P_k are known for each district then the a_k can be estimated by regressing the variable $M_k/\overline{M}_k$ on P_k. If the M_k are not known but the $\overline{M}_k$ are available, then the a_k can be estimated by regressing $M/\overline{M}$ on the variables $W_k P_k$ in a joint regression. If the W_k are not known then the regression of $M/\overline{M}$ on the K pollution variables P_k will yield estimates of $W_k a_k$. This will permit an estimate of $\overline{\Delta M}/\overline{M}$ or the proportional effect on mortality. It will not permit an estimate of the weighed $\bar{P}$ and hence of the weighed $\bar{a}$ characterizing the whole area.

Incomplete set of measuring stations. A number of simple cases will be examined to show how the deletion of measuring stations from the set will affect the total estimate of effects. If the correlation coefficient of a particular P_k is always zero when correlated with any other P_k then its omission from the multivariate regression will reduce the estimate of total effect, assuming the b_k are positive. Thus, under the assumption that P_k variables are not correlated with each other and the b_k are positive, the omission of a subset of the P_k should tend always to yield an estimate that is too low.

Consider an example at the opposite extreme, where the correlation between all pairs of P_k is one. Then any P_k variable will give a proper estimate of total effect. The weighed $\bar{a}$ cannot be estimated in the absence of the $\overline{P}_k$ and their appropriate weights.

If it is assumed that the set of P_k values can be divided into two subsets and for each subset the correlation coefficients are one but the correlation between two variables from different subsets is minus one, then it is readily seen that an estimate based on a single P_k will yield the difference

of effects in the two subareas corresponding to the two subsets P_k rather than the total mortality effect.

Similarly, it can be seen that if all the P_k can be expressed as a linear sum with positive coefficients of a subset then the availability of the subset and their use in a multiple regression will give a proper estimate of $\overline{\Delta M}/\overline{M}$. Where the available subset does not fully define the P_k but provides a linear form with positive coefficients plus an uncorrelated component, then only part of the effect will be measured. This appears to be the case for New York City and it will be interesting to attempt to express a number of stations as a linear form of the laboratory station and Roosevelt Island station measurements.

Finally, if in the above consideration some coefficients of the linear forms are negative, then the situation will resemble the earlier one where correlation was either +1 or -1, and improper estimates will be obtained for either total or partial effects.

The above illustrations may be generalized readily to the case of more than one type of pollutant measurement. Whether dealing with one type or several pollutants, it will be seen that generally the absence of a full set of P_k will result in a reduced estimate.

REFERENCES

1. Schimmel, H. and Murawski, T. H.: The relation of pollution to mortality. *J. Occup. Med. 18*:316-33, 1976.
2. Schimmel, H. and Greenburg, L.: A study of the relation of pollution to mortality—New York City, 1963-1968. *J. Air Pollut. Control Assoc. 22*:616, 1972.
3. Greenburg, L. et al.: Report of an air pollution incident in New York City, November 1953. *Public Health Rep. 77*:7-16, 1962; Greenburg, L. et al.: Intermittent air pollution episode in New York City, 1962. *Public Health Rep. 78*:1061-64, 1963; Greenburg, L. et al.: Air pollution, influenza, and mortality in New York City. *Arch. Environ. Health 15*:430-38, 1967; Glasser, M., Greenburg, L., and Field, F.: Mortality and morbidity during a period of high levels of air pollution, New York, Nov. 23 to 25, 1966. *Arch. Environ. Health 15*:684-94, 1967; Glasser, M. and Greenburg, L.: Air pollution, mortality and weather. *Arch. Environ. Health 22*:334-43, 1971.
4. Buechley, R. W. et al.: SO_2 levels and perturbations in mortality. A study in the New York-New Jersey metropolis. *Arch. Environ. Health 27*:134-37, 1973.
5. Mazumdar, S. and Sussman, N.: Personal communication.
6. Hearings before the New York State Department of Environmental Conservation and New York City Department of Public Health on a request by Con Ed to permit burning of higher sulfur fuel, 1977.
7. Susser, M.: *Causal Thinking in the Health Sciences: Concepts and Strategies in Epidemiology*. New York, Oxford University Press, 1973.
8. Buechley, R. W.: Discussion of paper by Herbert Schimmel. *Bull. N.Y. Acad. Med. 54:*1113-14, 1978.
9. Goldstein, I. F., Landovitz, L., and Block, G.: Air pollution patterns in New York City. *J. Air Pollut. Control Assoc. 24*:148-52, 1974.