United States Environmental Protection Agency
Health Effects Research Laboratory
Research Triangle Park NC 27711
EPA-600/1-78-043
June 1978

Research and Development

EPA

Comparison of Methods for the Analysis of Panel Studies

RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, U.S. Environmental Protection Agency, have been grouped into nine series. These nine broad categories were established to facilitate further development and application of environmental technology. Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields. The nine series are:

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies
6. Scientific and Technical Assessment Reports (STAR)
7. Interagency Energy-Environment Research and Development
8. "Special" Reports
9. Miscellaneous Reports

This report has been assigned to the ENVIRONMENTAL HEALTH EFFECTS RESEARCH series. This series describes projects and studies relating to the tolerances of man for unhealthful substances or conditions This work is generally assessed from a medical viewpoint, including physiological or psychological studies. In addition to toxicology and other medical specialities, study areas include biomedical instrumentation and health research techniques utilizing animals — but always with intended application to human health measures.

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

EPA-600/1-78-043
June 1978

COMPARISON OF METHODS FOR THE ANALYSIS OF PANEL STUDIES

by

Victor Hasselblad
Statistics and Data Management Office
Health Effects Research Laboratory
Research Triangle Park, N.C. 27711

U.S. ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF RESEARCH AND DEVELPMENT
HEALTH EFFECTS RESEARCH LABORATORY
RESEARCH TRIANGLE PARK, N.C. 27711

DISCLAIMER

FOREWORD

The many benefits of our modern, developing, industrial society are accompanied by certain hazards. Careful assessment of the relative risk of existing and new man-made environmental hazards is necessary for the establishment of sound regulatory policy. These regulations serve to enhance the quality of our environment in order to promote the public health and welfare and the productive capacity of our Nation's population.

The Health Effects Research Laboratory, Research Triangle Park, conducts a coordinated environmental health research program in toxicology, epidemiology, and clinical studies using human volunteer subjects. These studies address problems in air pollution, non-ionizing radiation, environmental carcinogenesis and the toxicology of pesticides as well as other chemical pollutants. The Laboratory participates in the development and revision of air quality criteria documents on pollutants for which national ambient air quality standards exist or are proposed, provides the data for registration of new pesticides or proposed suspension of those already in use, conducts research on hazardous and toxic materials, and is primarily responsible for providing the health basis for non-ionizing radiation standards. Direct support to the regulatory function of the Agency is provided in the form of expert testimony and preparation of affidavits as well as expert advice to the Administrator to assure the adequacy of health care and surveillance of persons having suffered imminent and substantial endangerment of their health.

This report addresses the extreme difficulty of analyzing panel data collected over time in an epidemiologic study. Although no exact solution is found, some improvements are suggested for the design and analysis of future studies.

F. G. Hueter, Ph. D.
Acting Director,
Health Effects Research Laboratory

ABSTRACT

Three different methods of analysis of panels were compared using asthma panel data from a 1970-1971 study done by EPA in Riverhead, N. Y. The methods were 1) regression analysis using raw attack rates; 2) regression analysis using the ratio of observed attacks to expected attacks; and 3) discriminant analysis where repeated attacks were ignored. The first two methods were found to have serious serial correlation problems. The third method eliminated this problem, but reduced the effective sample size considerably.

A more appropriate method was suggested for larger panels over shorter periods of time. The analyses of the Riverhead data showed that any sulfate effect on asthmatics was confounded with seasonal trends.

CONTENTS

FIGURES

TABLES

ACKNOWLEDGMENT

I would like to thank the three reviewers, John Creason, Larry Kupper, and Robert Chapman, who all made excellent suggestions. I would also like to thank William C. Nelson for his continued support and suggestions, and Barbara Crabtree for the typing of this manuscript.

SECTION 1

INTRODUCTION

Panel studies have made an important contribution to epidemiology. These include studies of asthmatics by Cohen, et al.[1], Salvaggio, et al.[2], and Girsh, et al.[3]. Lawther, et al.[4] looked at panels of bronchitics. Goldberg, et al.[5] looked at cardiovascular symptoms in elderly persons. Hammer, et al.[6], studied irritation symptoms, including eye discomfort. The effect of meteorology and pollution on pulmonary function was studied by Lawther, et al.[7,8]. The problem with panel studies, as noted by Stebbings and Hayes[9], is that the analysis of such studies is statistically difficult. If the data are continuous, such as pulmonary function data, then a repeated measures analysis or multivariate analysis of variance is appropriate. If the data are discrete, such as yes-no data, then no available techniques are fully satisfactory.

In an effort to evaluate a range of analysis techniques, this paper presents a comparison of three different methods of analysis, using 1970-1971 Riverhead, N. Y. asthma panel data. An analysis of this data, along with similar analyses of panels from Queens and Bronx was previously presented by Finklea, et al.[10] This study was one of several conducted by the Environmental Protection Agency's Community Health Environmental Surveillance System (CHESS).

SECTION 2

THE BASIC DATA

Panelists were selected from three New York communities. Riverhead, located near the center of Long Island, had generally lower exposures to air pollution than did the communities of Queens and Bronx. The names of panelists were obtained from hospital clinic records and records of practicing physicians. Background information on each panelist, including age and smoking history, was obtained by an interview. All panelists lived within a 1.5-mile radius of a monitoring station.

From October 4, 1970 to May 29, 1971 each panelist received a weekly diary. The panelist recorded the presence or absence of an asthma attack for each day of the week, and then returned the completed diary by mail. During this same time period, continuous 24-hour air monitoring was also conducted by EPA. The measurements included total suspended particulates (TSP), sulfate fraction of TSP (SO_x), nitrate fraction of TSP (NO_x), sulfur dioxide (SO_2), and nitrogen dioxide (NO_2). The measurement methods are described by Finklea, et al.[10].

This paper will restrict itself to the analysis of the Riverhead panel. Although Riverhead had generally lower pollution values than the other two communities, the previous analysis of the Riverhead data showed a significant relationship between asthma attack rate and SO_x. All of the pollutants are correlated with one another, as shown in Table 1. This paper will consider the relationship of asthma with SO_x only.

The first 2 weeks of data were not used, to minimize the problem of initial over-reporting, and to allow the panelists to become familiar with the study. Fifty panelists were initially enrolled in Riverhead, but 11

TABLE 1. PAIRWISE CORRELATIONS OF POLLUTANTS IN RIVERHEAD, N. Y. FOR THE TIME PERIOD OCTOBER 18, 1970 TO MAY 29, 1971.*

	TSP	SO_X	NO_X	SO_2
TSP	1.000 (206)	.796 (205)	.487 (206)	.370 (188)
SO_X	.796 (205)	1.000 (205)	.380 (205)	.377 (187)
NO_X	.487 (206)	.380 (205)	1.000 (206)	.141 (188)
SO_2	.370 (188)	.377 (187)	.141 (188)	1.000 (194)

* Sample Sizes Are Given in Parentheses.

were eliminated from the analyses because they either never returned a diary, or never reported having an asthma attack at any time during the study.

SECTION 3

ANALYSES USING THE RAW ATTACK RATE

One obvious method of analysis is to use the raw attack rate as the dependent variable. This rate is defined for each day as the number of panelists reporting an attack on that date divided by the number of panelists giving a response on that same day. Figure 1 shows a histogram of this variable. The analyses of this variable were restricted to the 205 days with valid SO_x measurements. The attack rate ranged from .057 to .500, with a mean value of .181. The SO_x measurements ranged from .5 to 51.4, with a mean of 9.81. A histogram of the SO_x values is shown in Figure 2. Note that both the simple attack rate and SO_x have distributions which are skewed to the right.

There are a large number of possible covariates which could be used in the analyses. Several of these, such as pollen or emotional stress, were not measured in this study. The following factors were used in the analyses shown in Table 2.

1) Season: several authors have noted seasonal trends in asthma, with higher rates in October and November in the northern hemisphere[3,11-13] and higher rates in April and May in the southern hemisphere[14]. This effect was estimated using the first two terms of a Fourier series.[15]

2) Start-up effect: there is often over-reporting during the first few weeks of a study. Although the first 2 weeks were dropped, the reciprocal of the week number was still used as a covariate.

3) Temperature: low temperatures have also been associated with increased asthma attacks.[12-14,16] The daily minimum temperature as reported at the airport was included as a covariate.

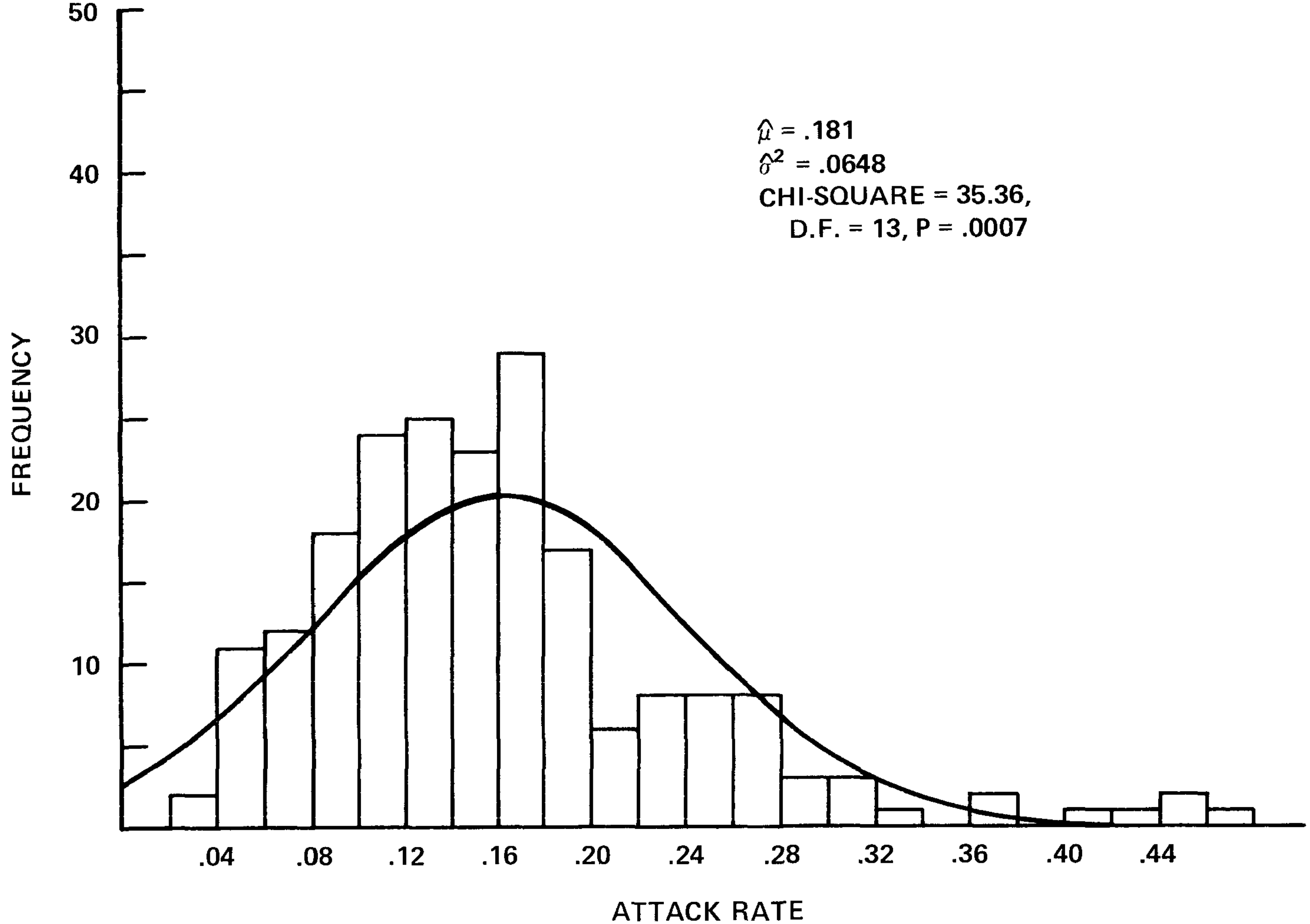

Figure 1. Histogram of the attack rate for the 1970-1971 Riverhead asthma panel. A normal distribution is fitted to the data.

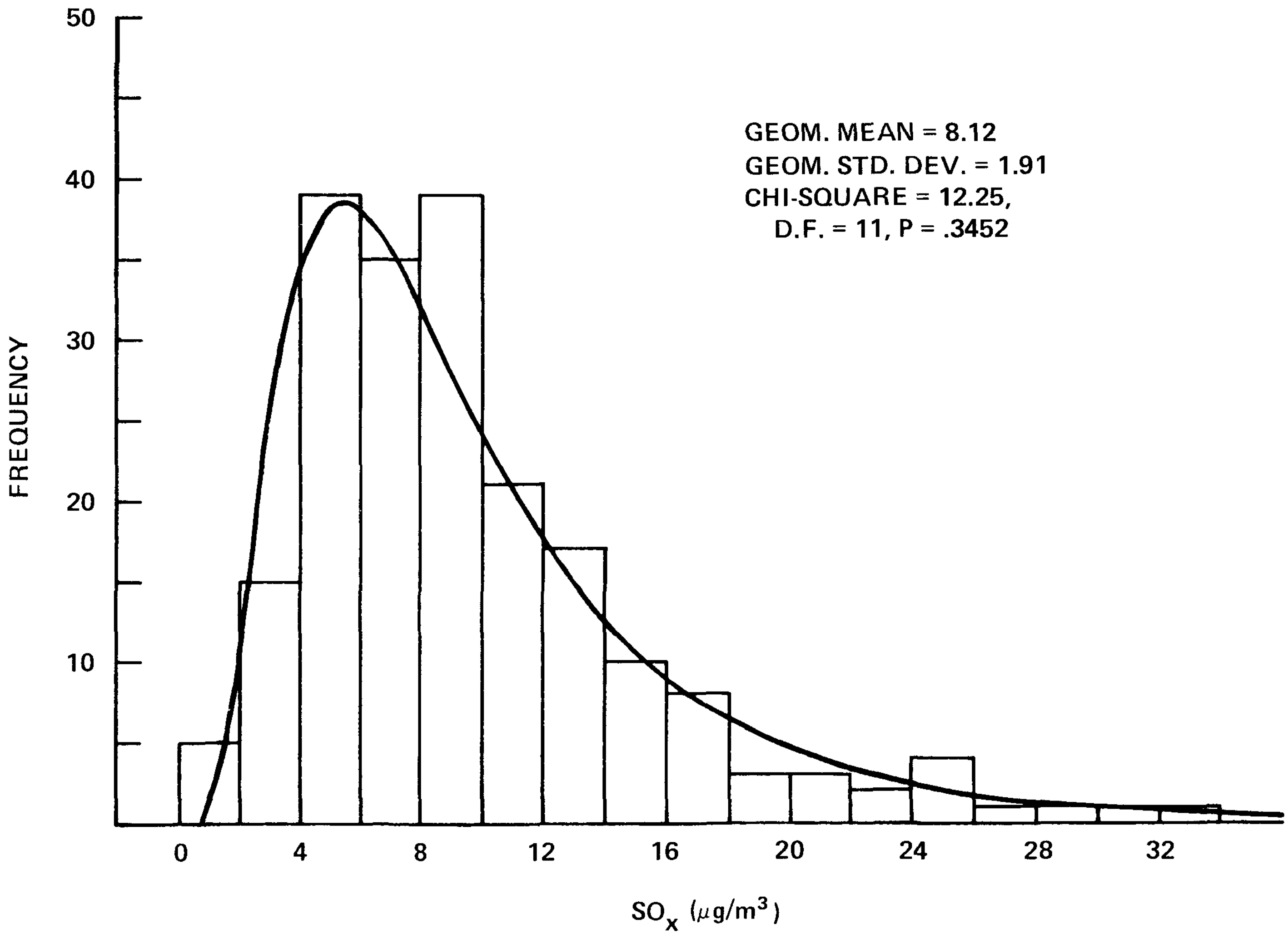

Figure 2. Histogram of suspended sulfate values (SO_x) for the 1970-1971 Riverhead asthma panel. A lognormal distribution is fitted to the data.

TABLE 2. MULTIPLE REGRESSION ANALYSES OF THE RIVERHEAD ASTHMA ATTACK RATE ON SO_x USING VARIOUS COVARIATES SEPARATELY

Covariate	Degrees of Freedom	F-to-enter for Covariate	SO_x Regression Coefficient	F-to-remove for SO_x	P Value
(None)	-	-	.00188	4.87	.0285
Season	4	42.30	.00049	.48	.4877
Start-up	1	105.24	.00140	4.67	.0318
Minimum Temperature	1	.22	.00194	5.04	.0259
Day of Week	6	1.06	.00196	5.25	.0230

4) Day of week: this variable was not known to be important, but dummy variables were included to allow for any possible association of attack with day of week.

Table 2 shows the F-to-enter for all of these covariates. This is the F value associated with a simple linear regression using the particular covariate as the independent variable. The SO_x regression coefficient is given for each covariate added separately to the model. The SO_x F-to-remove value, which is a test of the effect of SO_x in addition to that of a specified covariate, is also shown. Although both the seasonal trend and the start-up effect are highly significant, only the seasonal trend reduces the significance of SO_x.

Table 3 shows the result of including all covariates in the same multiple regression analysis. The analysis demonstrates the effect of too many correlated covariates. The overall sum of squares explained is almost 10 times the sum of the individual sum of squares explained. The regression coefficient for the start-up effect is negative, even though it has a positive simple correlation of .653.

The residuals based on the regression analysis in Table 3 are shown in Figure 3. A chi-square goodness of fit test for normality of the residuals gives a value of 13.28 for 9 degrees of freedom (p=.1503). There is some evidence of skewness to the right (p=.0100), but less evidence of Kurtosis (p=.0738), using Fisher's G statistics[17]. The greatest problem with the residuals is that the serial correlation is .3335, which is highly significant (p=<.0001).

There are at least three major problems with the analysis:

1) There is no adjustment for changing panel composition due to

dropins and dropouts.

2) The information about the sample size of the panel is lost.

3) The residuals show high serial correlation.

TABLE 3. MULTIPLE REGRESSION ANALYSIS OF THE RIVERHEAD ASTHMA ATTACK RATES ON SO_x USING ALL COVARIATES SIMULTANEOUSLY

Covariate	Degrees of Freedom	Regression Coefficient	Sum of Squares	Mean Square	F-to Remove	P Value
SO_x	1	.00078	.00408	.00408	1.18	.2780
Season	4	-	.03670	.00918	2.66	.0340
Start-up	1	-.03919	.00006	.00006	.02	.8912
Minimum Temperature	1	-.00102	.00408	.00408	2.30	.1311
Day of Week	6	-	.03468	.00578	1.68	.1286
Overall	13	-	.66956	.05150	14.94	<.0001
Error	191		.65826	.00345		

Multiple R^2 = .50425

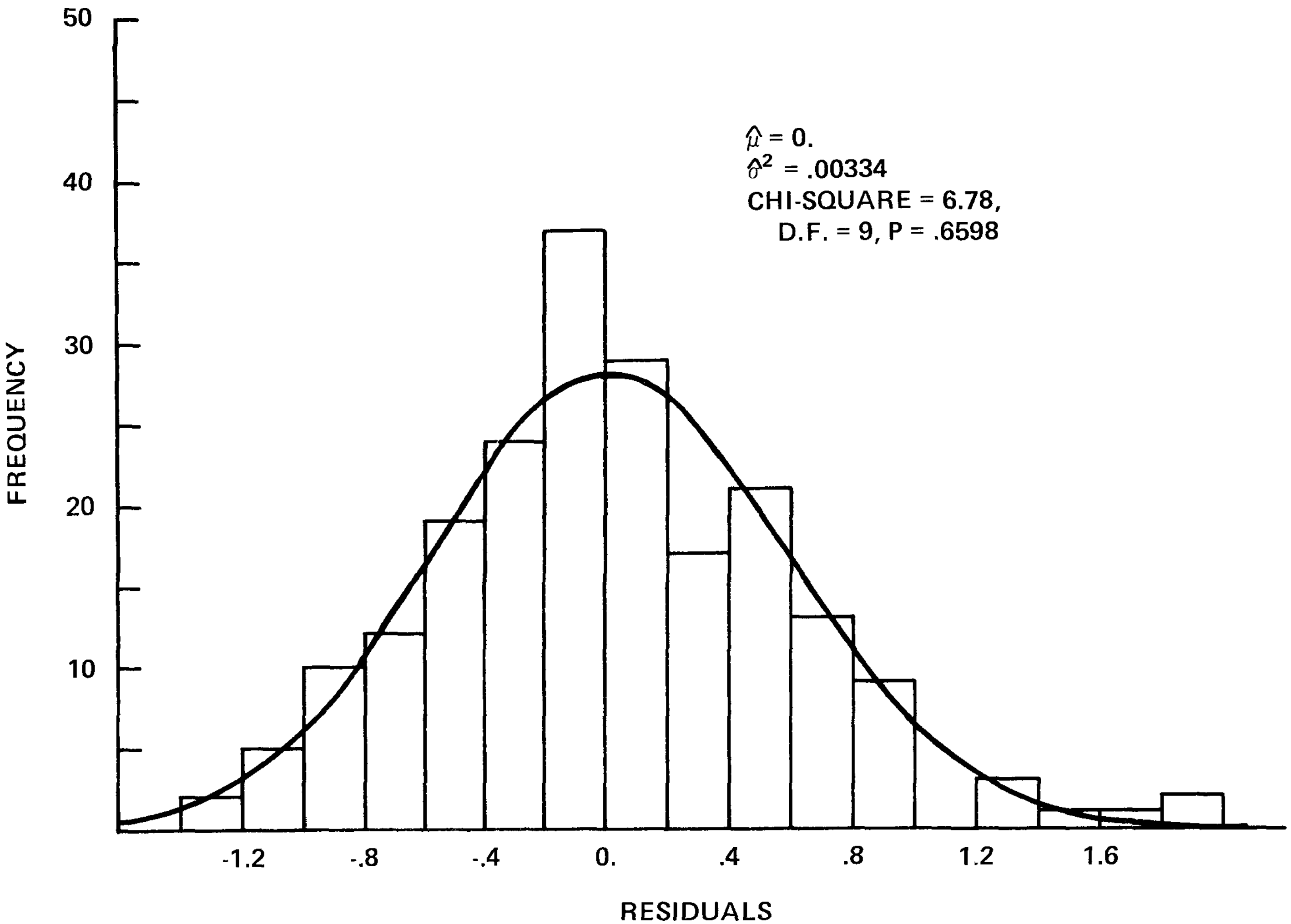

Figure 3. Histogram of residuals from a regression analysis of attack rate for the 1970-1971 Riverhead asthma panel. A normal distribution is fitted to the data.

SECTION 4

ANALYSES USING OBSERVED/EXPECTED

One refinement of the analysis in the previous section is to compute an expected attack rate for each day. This is the sum of the mean daily attack rates for each individual who responded on a particular day. By forming the ratio of the observed to the expected for each day, the analysis can allow for changing panel composition. This ratio ranged from .3378 to 2.3861. A histogram of the ratio is shown in Figure 4.

Table 4 was constructed in the same manner as Table 2, except that the ratio of the observed to expected was used as a dependent variable instead of the raw attack rate. The results are quite similar to Table 2, with season again eliminating any SO_x effect. The p values for SO_x are slightly larger than they were for the analysis in Table 2.

If all covariates are added simultaneously, as was done in Table 3, the result is as shown in Table 5. The covariates are still highly correlated, and as a result, the individual sums of squares do not add up to one fifth of the overall sum of squares. Again, the regression coefficient for the start-up effect is negative, even though it is positively correlated (r=.562) with the ratio of observed to expected rates.

The residuals based on the regression analysis in Table 5 are shown in Figure 5. A chi-square goodness of fit test for normality of the residuals gives a value of 24.96 for 21 degrees of freedom (p=.2488). There is less evidence of skewness (p=.0916) and kurtosis (p=.9422), than in the previous analysis. The residuals still have a significant serial correlation of .2257 (p=.0010).

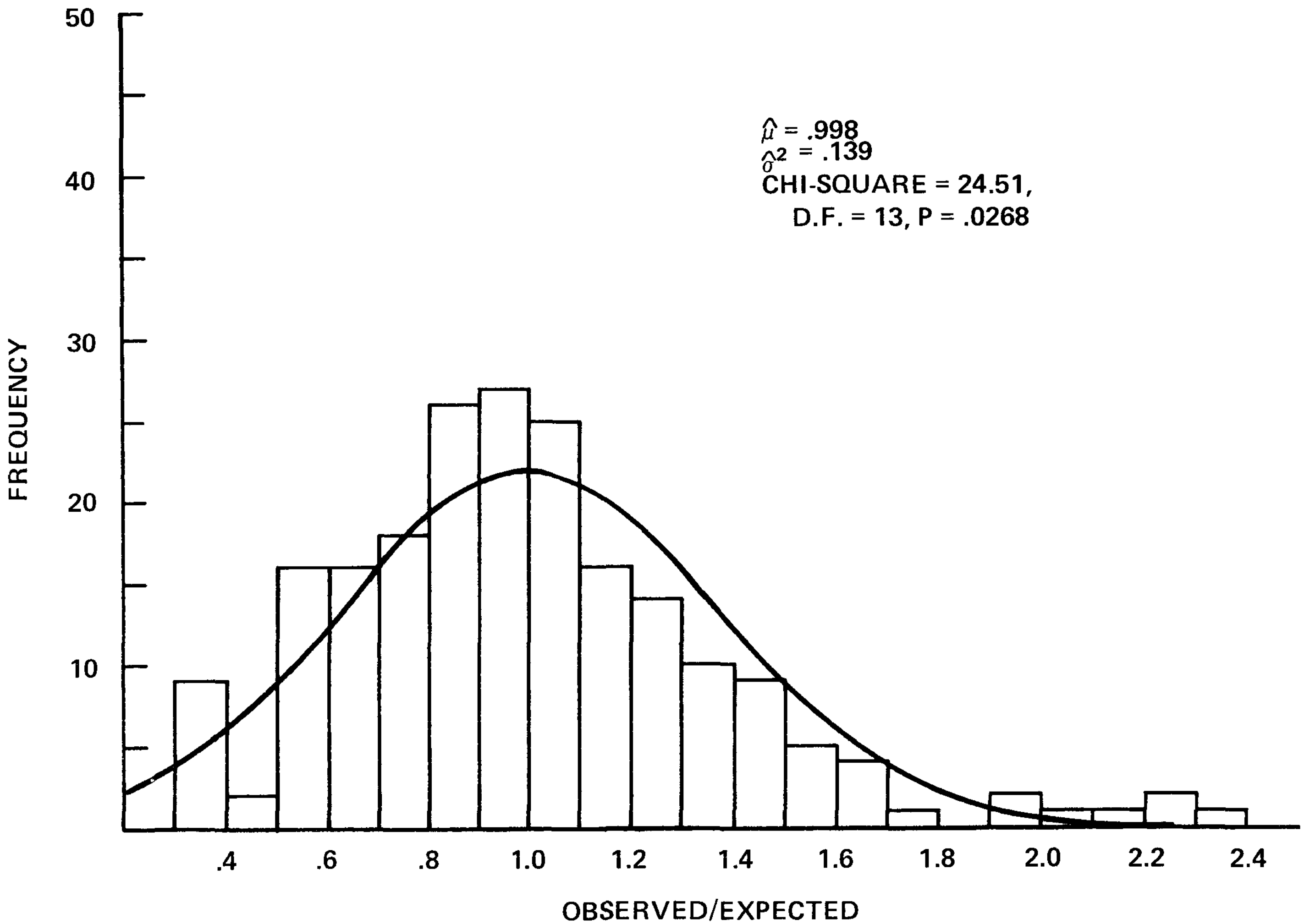

Figure 4. Histogram of the observed attacks divided by the expected attacks for the 1970-1971 Riverhead asthma panel. A normal distribution is fitted to the data.

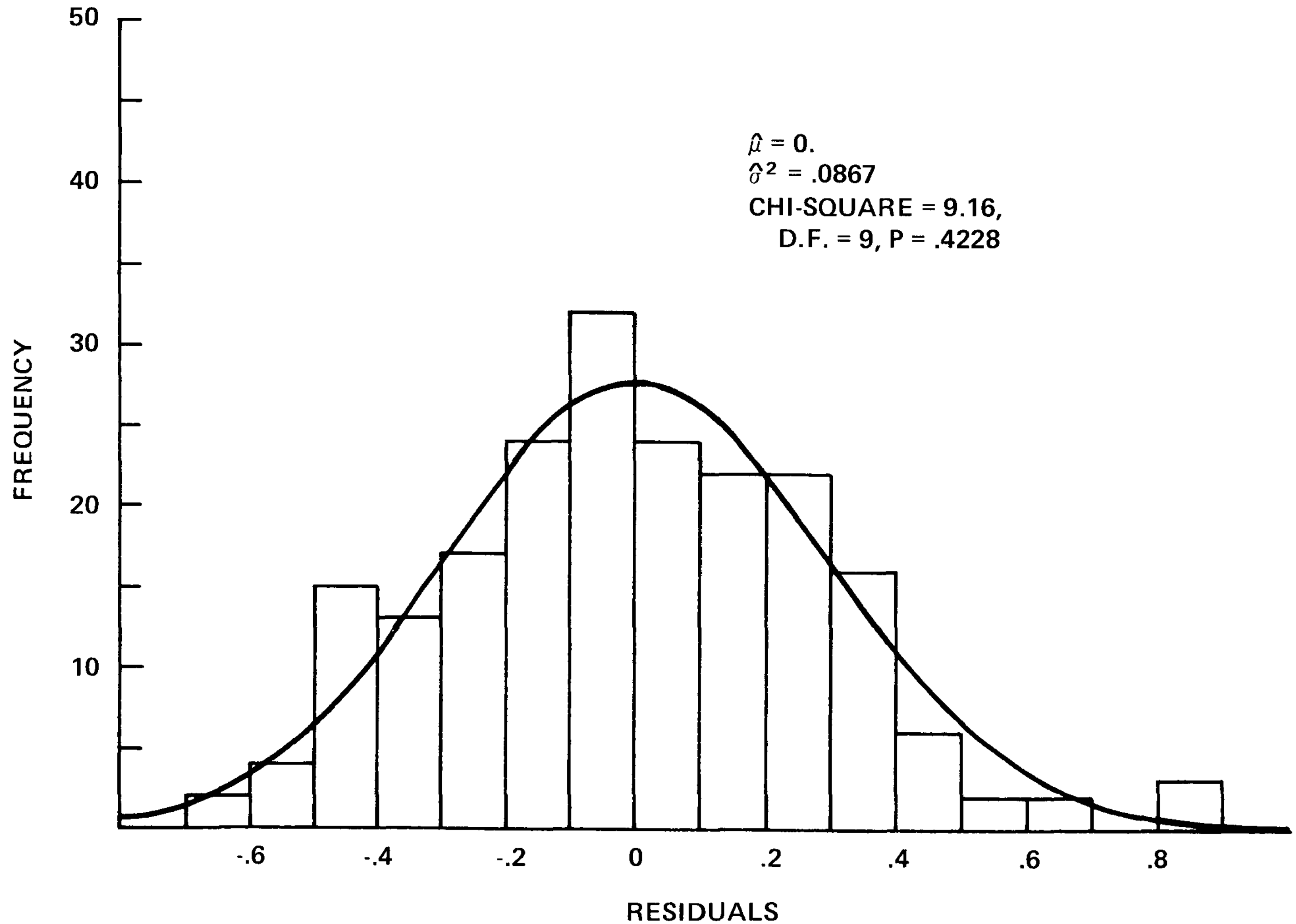

Figure 5. Histogram of residuals from a regression analysis of observed/ expected for the 1970-1971 Riverhead asthma panel. A normal distribution is fitted to the data.

TABLE 4. MULTIPLE REGRESSION ANALYSES OF THE RATIO OF OBSERVED TO EXPECTED ON SO_x USING VARIOUS COVARIATES SEPARATELY

Covariate	Degrees of Freedom	F-to-enter for Covariate	SO_x Regression Coefficient	F-to-remove for SOx	P Value
(None)	-	-	.00801	4.14	.0432
Season	4	26.87	.00279	.61	.4345
Start-up	1	93.50	.00622	3.62	.0587
Minimum Temperature	1	.02	.00793	3.95	.0481
Day of Week	6	1.47	.00842	4.59	.0333

TABLE 5. MULTIPLE REGRESSION ANALYSIS OF THE RATIO OF OBSERVED TO EXPECTED ON SO_x USING ALL COVARIATES SIMULTANEOUSLY

Covariate	Degrees of Freedom	Regression Coefficient	Sum of Squares	Mean Square	F-to Remove	P Value
SO_x	1	.00457	.13823	.13823	1.56	.2128
Season	4	-	.66025	.16506	1.87	.1180
Start-up	1	-.56467	.01342	.01342	.15	.6974
Minimum Temperature	1	-.00535	.21718	.21718	2.46	.1188
Day of Week	6	-	1.13355	.18892	2.14	.0511
Overall	13	-	11.65512	.89655	10.14	<.0001
Error	191		16.8949	.08845		

Multiple R^2 = .40825

SECTION 5

A MODIFIED DISCRIMINANT ANALYSIS

Any sophisticated analysis of panel data should consider the pattern of individual responses. The following method is a crude attempt at this. Each member of a panel is assumed to respond either yes or no to a particular question for each day of a particular panel study. Without loss of generality, we will assign the value 1 to yes, and 0 to no. The nonindependence of individual responses can be demonstrated easily. For each day, j, each person responds 0 or 1, as he does on day j + 1. This process is a natural Markov chain:

Day j + 1 Response

Day j Response	0	1
0	p_{00}	p_{01}
1	p_{10}	p_{11}

where $p_{00} + p_{01} = 1$, $p_{10} + p_{11} = 1$. The p's represent the probabilities of having a particular response on each of 2 successive days. If each person's response is independent of his response on the previous day, then p_{01} will equal p_{11}. This can be tested using the standard 2 x 2 chi-square test.

The independence, or lack thereof, was studied using the 1970-1971 Riverhead asthma data. The counts of responses were done for each individual, and recorded in the 2 x 2 table previously described. Let n_{00}, n_{01}, n_{10}, n_{11} be the observed counts corresponding to p_{00}, p_{01}, p_{10}, p_{11} respectively. The chi-square test for independence of one day's outcome from the previous day's is

$$\chi_1^2 = \frac{N(n_{00}n_{11}-n_{10}n_{01})^2}{(n_{00}+n_{01})(n_{00}+n_{10})(n_{10}+n_{11})(n_{01}+n_{11})}$$

where $N = n_{00}+n_{01}+n_{10}+n_{11}$

These were computed for each panelist having marginal totals greater than or equal to 10. Of the 21 panelists meeting this criteria, results for 17 were significant at the .01 level, 1 was significant at a level between .01 and .05, and 3 were not significant at a level less than .05.

Another interesting statistic is the relative risk.[18] This is the probability of developing an attack on one day given at attack on the previous day, divided by the probability of developing one given none on the previous day. The observed relative risks ranged from 1.07 to 37.55. The median relative risk was 4.96. Thus the "median" person was almost five times as likely to have an attack if he had one the previous day. The assumption of independence is clearly not met.

One way to avoid the problem of lack of independence is to look only at p_{00} and p_{01}. Let the analysis explain why a person remains in state 0 (no symptoms) on some days, and why he moves from state 0 to state 1 (develops symptoms) on other days. The price of this restriction is to discard two other types of information: 1) p_{10}: the person moves from state 1 to state 0 (loses symptoms) and 2) p_{11}: the person stays in state 1 (retains symptoms). It could be argued that these two events are more dependent on a person's recovery mechanism than on the insult causing the response.

Using p_{00} and p_{01} only, we can define a variable y_{ij} as follows:

y_{ij} = 0 if person i has no symptoms on either day j or day j-1

= 1 if person i has symptoms on day j, but none on day j=1

= undefined otherwise.

y_{ij} can be thought of as a Bernoulli trial, with p_{ij} given by

$$p_{ij} = h(\mu_i, x_{1j}, \ldots, x_{kj})$$

where μ_i = overall symptom rate for person i and $x_{1j}, \ldots, x_{kj}$ = the k values of the independent variables for day j. These effects could include such factors as pollution, meteorology, seasonal cycles, and day of week effects. The logical functional form would be a general dose-response curve of the form

$$p_{ij} = \int_{-\infty}^{g(\mu_i, x_{ij} \ldots, x_{ky})} f(t)\, dt$$

where f(t) is either a probit or logit density. Even if g is a linear function, the large number of variables makes maximum likelihood or least squares estimation of the coefficients extremely difficult.

If, however, the effects of the daily factors are relatively small, then the dose-response curve may be approximately linear in those variables. If we assume that p_{ij} is linear in the parameters to be estimated, we have reduced the problem to a point that some analyses can be calculated.

The solution is to run a discriminant analysis using the y_{ij}'s as dependent variables, and dummy variables for people along with the other covariates as independent variables. Approximate tests for significance for the independent variables can be made using partial F's, as described by Lachenbruch.[19] The results of these analyses appear in Tables 6 and 7. Table 6 is analogous to Tables 2 and 4, in that the covariates are added one

TABLE 6. DISCRIMINANT ANALYSES OF THE RIVERHEAD ASTHMA ATTACK RATE ON SO_x USING VARIOUS COVARIATES SEPARATELY

Covariate	Degrees of Freedom	F-to-enter for Covariate	SO_x Regression Coefficient	F-to-remove for SOx	P Value
(None)	-	-	.00093	2.77	.0961
Season	4	7.82	.00031	.28	.5972
Start-up	1	8.81	.00097	3.03	.0819
Minimum Temperature	1	6.80	.00079	2.02	.1557
Day of Week	6	3.01	.00098	3.07	.0798

TABLE 7. DISCRIMINANT ANALYSIS OF THE RIVERHEAD ASTHMA ATTACK RATES ON SO_x USING ALL COVARIATES SIMULTANEOUSLY

Covariate	Degrees of Freedom	Discriminant Coefficient	Sum of Squares	Mean Square	F-to Remove	P Value
SO_x	1	.00039	.02872	.02872	.43	.5098
Season	4	-	.98623	.24656	3.73	.0049
Start-up	1	-.08660	.13781	.13781	2.08	.1489
Minimum Temperature	1	-.00038	.03299	.03299	.50	.4800
Day of Week	6	-	1.28332	.21389	3.24	.0036
People	31	-	49.31891	1.59093	-	-
Error	4556		301.19530	.06611		

at a time. SO_x is no longer significant at the .05 level (p=.096) even without any covariates added. If a season effect is added, there again is no evidence of any SO_x effect. Table 7 is analogous to Tables 3 and 5, with all covariates added simultaneously. The discriminant analysis shows a stronger seasonal effect than did the other two analyses.

SECTION 6

COMPARISON OF METHODS

The discriminant analysis method has eliminated the three major problems of regression analysis on attack rates, namely (1) changing panel composition, (2) loss of sample size information, and (3) serial correlation of responses. In their place are some new problems:

1) A large number of positive responses are ignored (approximately 1/2 in this particular analysis.)
2) The true distributions of the "F to remove" tests are not known when discrete variables are used in discriminant analyses.
3) The procedure appears to be very conservative as a test for the effect of air pollution on asthma attack rates.

There are obviously several other possible methods of analysis. These include the "multiple logistic function" of Truett, Cornfield, and Kannel[20]; the "stimulus response" method of Lebowitz[21], and the "binary multiple regression analysis" of Elwood, Mackenzie, and Cran[22]. These methods are all appropriate for some data sets, but none of them solve the problems of this particular kind of panel study.

Perhaps a promising method of analysis is that given by Koch, et al.[23] This technique provides for the analysis of multivariate categorical data which are obtained from a repeated measure design. Unfortunately, the method requires that the number of days be much less than the number of subjects. For the asthma panels studied thus far by EPA, this limitation would mean that no more than one month's data could be used at one time. This deletion would further reduce the number of eligible panelists, since many have no attacks in any given month. The reduced data set would be insufficient for analysis.

SECTION 7

DISCUSSION

Although this report does not find a fully satisfactory method of analysis for asthma panel studies, a few conclusions can be drawn. The first is that any method of analysis must allow for the nonindependence of responses for a particular individual. Unfortunately, the few available methods which accomplish this reduce the data set considerably, so that consistent significant results are unlikely. It is possible, of course, that future research may provide a more appropriate technique. Until the time that better methods are available, a possible improvement is to have larger panels for shorter periods of time. These data could be analyzed by the method of Koch, et al.[23], and the contribution of seasonality factors would also be greatly reduced.

The results of the various analyses on the Riverhead panel indicate that the relation between SO_x and the asthma rate is confounded with the seasonal trends of both variables. Any statement of a positive relationship between the two variables must be made with much qualification.

REFERENCES

1. Cohen, A. A., S. Bromberg, R. W. Buechley, L. T. Heiderscheit, and C. M. Shy. Asthma and Air Pollution from a Coal-fueled Power Plant. Am. J. of Public Health, 62: 1181-1188, 1972.

2. Salvaggio, John, Victor Hasselblad, and L. T. Heiderscheit. New Orleans Asthma. II. Relationship of Climatologic and Seasonal Factors to Outbreaks, J. of Allergy, 45: 257-265, 1970.

3. Girsch, L. S., E. Shubin, C. Dick, and F. A. Schulaner. A Study on the Epidemiology of Asthma in Children in Philadelphia. J. of Allergy, 39: 347-357, 1967.

4. Lawther, P. J., R. E. Waller, and M. Henderson. Air Pollution and Exacerbations of Bronchitis. Thorax: 25: 525-539, 1970.

5. Goldberg, H. E., A. A. Cohen, J. F. Finklea, J. H. Farmer, F. B. Benson, and G. J. Love. Frequency and Severity of Cardiopulmonary Symptoms in Adult Panels: 1970-1971 New York Studies. In: Health Consequences of Sulfur Oxides: A Report from CHESS, 1970-1971. EPA-650/1-74-004, U. S. Environmental Protection Agency, Research Triangle Park, N. C., 1974.

6. Hammer, D. I., V. Hasselblad, B. Portnoy, and P. F. Wehrle. Los Angeles Student Nurse Study. Daily Symptom Reporting and Photochemical Oxidants. Archives of Environmental Health, 28: 255-260, 1974.

7. Lawther, P. J., A. G. F. Brooks, P. W. Lord, and R. E. Waller. Day-to-day Changes in Ventilatory Function in Relation to the Environment. Part I. Spirometric Values. Environmental Research, 7: 27-40, 1974.

8. Lawther, P. J., A. G. F. Brooks, P. W. Lord, and R. E. Waller. Day-to-day Changes in Ventilatory Function in Relation to the Environment. Part II. Peak Expiratory Flow Values. Environmental Research, 7: 41-53.

9. Stebbings, James H., Jr., and Carl G. Hayes. Panel Studies of Acute Health Effects of Air Pollution I. Cardiopulmonary Symptoms in Adults, New York 1971-1972. Environmental Research, 11: 89-111, 1976.

10. Finklea, John F., John H. Farmer, Gory J. Love, Dorothy C. Calafiore, and Wayne G. Sovocool. Aggravation of Asthma by Air Pollutants: 1970-1971 New York Studies. In: Health Consequences of Sulfur Oxides: A Report form CHESS, 1970-1971, EPA-650/1-74-004, U. S. Environmental Report from CHESS, 1970-1971, EPA-650/1-74-004, U. S. Environmental Protection Agency, Research Triangle Park, N. C., 1974.

11. Greenberg, L., and F. Field. Air Pollution and Asthma. J. of Asthma Research, 2: 195, 1965.

12. Booth, S., L. Degroot, R. Markash, and R. J. M. Horton. Detection of Asthma Epidemics in Seven Cities. Archives of Environmental Health, 10: 152-155, 1965.

13. Shy, Carl M., Victor Hasselblad, Leo T. Heiderscheit, and Arlan A. Cohen. Environmental Factors in Bronchial Asthma. In: Environmental Factors in Respiratory Disease, Douglas H. K. Lee, ed. Academic Press, 1972.

14. Derrick, E. H. The Annual Variation of Asthma in Brisbane: Its Relation to the Weather. International J. of Biometeorology, 10: 91-99, 1966.

15. Taylor, Angus E. Advanced Calculus. Ginn and Company, 1955. pp. 711-718.

16. Tromp, S. W. Biometeorological Analysis of the Frequency and Degree of Asthma Attacks in the Western Part of the Netherlands. In: Proceedings of the Second International Bioclimatological Conference. Oxford: Pergamon Press, 1962. p. 477.

17. Fisher, Ronald A. Statistical Methods for Research Workers. Hafner, 1958. pp. 52-54.

18. Fleiss, Joseph L. Statistical Methods for Rates and Proportions. John Wiley & Sons, New York, 1973. pp. 39-51.

19. Lachenbruch, Peter A. Discriminant Analysis. Hafner Press, 1975. pp. 27-29.

20. Truett, Jeanne, Jerome Cornfield, and William Kannel. A Multivariate Analysis of the Risk of Coronary Heart Disease in Framingham. J. of Chronic Disease, 20: 511-524, 1967.

21. Lebowitz, Michael D. A Stimulus Response Method for the Analysis of Environmental and Health Events Related in Time. Environmental Letters, 2: 23-34, 1971.

22. Elwood, J. H., G. Mackenzie, and G. W. Cran. The Measurement and Comparison of Infant Mortality Risks by Binary Multiple Regression Analysis. J. of Chronic Disease, 24: 93-106, 1971.

23. Koch, Gary G., J. Richard Landis, Jean L. Freeman, Daniel H. Freeman, Jr., and Robert G. Lehnen. A General Method for the Analysis of Experiments with Repeated Measurement of Categorical Data. Biometrics, 33: 133-158, 1977.

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/1-78-043	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE: COMPARISON OF METHODS FOR THE ANALYSIS OF PANEL STUDIES	5. REPORT DATE: June 1978
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S): Victor Hasselblad	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Statistics and Data Management Office, Health Effects Research Laboratory, Research Triangle Park NC 27711	10. PROGRAM ELEMENT NO: 1AA601
	11. CONTRACT/GRANT NO
12. SPONSORING AGENCY NAME AND ADDRESS: Health Effects Research Laboratory RTP,NC; Office of Research and Development; U.S. Environmental Protection Agency; Research Triangle Park, NC 27711	13. TYPE OF REPORT AND PERIOD COVERED
	14. SPONSORING AGENCY CODE: EPA 600/11

15. SUPPLEMENTARY NOTES

16. ABSTRACT

Three different methods of analysis of panels were compared using asthma panel data from a 1970-1971 study done by EPA in Riverhead, New York. The methods were (1) regression analysis using raw attack rates; (2) regression analysis using the ratio of observed attacks to expected attacks; and (3) discriminant analysis where repeated attacks were ignored. The first two methods were found to have serious serial correlation problems. The third method eliminated this problem, but reduced the effective sample size considerably.

A more appropriate method was suggested for larger panels over shorter periods of time. The analyses of the Riverhead data showed that any sulfate effect on asthmatics was confounded with seasonal trends.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
statistical analysis epidemiology asthma		06 F 12 A

18. DISTRIBUTION STATEMENT: RELEASE TO PUBLIC	19. SECURITY CLASS (This Report): UNCLASSIFIED	21. NO. OF PAGES: 36
	20. SECURITY CLASS (This page): UNCLASSIFIED	22. PRICE

EPA Form 2220-1 (9-73)