Health Costs of Air Pollution: A Study of Hospitalization Costs

BEN H. CARPENTER, MS, CHE, JAMES R. CHROMY, PHD, WALTER D. BACH, PHD, D. A. LESOURD, PHD, AND DONALD G. GILLETTE, PHD

Abstract: This study of the hospitalization costs of exposure to air pollution in Allegheny County, Pennsylvania was conducted to determine whether persons exposed to air pollution incurred higher hospital utilization rates and additional costs for treatment. A hospitalization data-base comprising 37,818 total admissions for respiratory, suspect circulatory diseases, and comparison circulatory diseases was tested in a cross-section type analysis for relationships between rates of hospitalization, length of stay, and levels of air quality in the neighborhoods of patients' residence. Air quality was identified using data from 49 monitoring stations. Corrections were made for race, age, sex, smoking habits, median income, and occupation.

The results show that hospitalization rates, length of stay, and costs of respiratory and suspect circulatory system diseases were significantly greater among populations residing in the more polluted zones of the County. At average costs for hospitalization in this area in 1972, the total increased cost for the 1.6 million persons in the County was estimated at $9.8 million ($9.1 million for increased hospitalization rates and $0.7 million for increased length of stay). The total health costs resulting from air pollution exposure in this area would be much greater when non-hospitalization costs are also included. (*Am J Public Health* 69:1232–1241, 1979.)

Air pollution damage to human health has been investigated by Ridker,[1] Lave and Seskin,[2] and Park.[3] These and other studies have been reviewed and assessed by Waddell.[4] Sterling, et al, studied urban morbidity as related to daily air pollution.[5–7] Hospitalization costs, however, have not been estimated. Outpatient medical costs for treatment of respiratory diseases were studied by Jacksch and Stoevener,[8] who found that, while air pollution may have affected the frequency of outpatient visits, it appeared not to have affected costs per contact with the medical system.

This study was made to determine whether persons exposed to air pollution were hospitalized more often or incurred more costs incident to hospitalization than persons not so exposed. The objectives were to develop an air quality data base and a hospitalization data base which could be merged with population data for analysis; and to estimate by appropriate methods the effects of exposure to pollutants on rates of hospitalization, length of stay in the hospital, and associated costs. Three classes of diseases were studied: respiratory diseases, suspect circulatory diseases, and comparison circulatory diseases.

Address reprint requests to Ben H. Carpenter, MS, ChE, Research Triangle Institute, P.O. Box 12194, Research Triangle Park, NC 27709. The author's academic affiliations are: Mr. Carpenter, West Virginia University; Dr. Chromy, statistician, North Carolina State University; Dr. Bach, meteorologist, University of Oklahoma; Dr. LeSourd, economist, Pennsylvania State University; and Dr. Gillette, economist, University of Maryland. This paper, submitted to the Journal June 12, 1978, was revised and accepted for publication June 29, 1979.

Editor's Note: See also related editorial, page 1210 this issue.

Methods

Technical Approach

A general concept of the excess risk of health costs because of exposure to pollutants was defined by the equation:

$$\text{Health Costs} = \sum_{ij} p_i c_{ij} r_{ij}$$

where p_i is the number of persons in the subpopulation exposed to pollutant level i, c_{ij} is the type j cost to a person in subpopulation i, and r_{ij} is the extra risk to subpopulation i of incurring the cost of type j. Each term of the health cost equation accounts for the added health costs of selected illnesses when a specified subpopulation is exposed to air pollution at a specified level.

The exposure levels, i, may be defined as the separate levels of each pollutant or as the combination of levels of several pollutants to which subpopulations are exposed. Either current air quality standards or other candidate values might be used to specify clean air if the relationships between air quality, c, and r proved to be sufficiently firm.

In this study, only hospitalization costs were estimated, but other types of costs can be accommodated by the health costs model.

Before these costs elements could be estimated, it was necessary to establish that c and r were related to the levels of pollutants. The conceptual health costs model had to be verified; verification of the model required suitable data to test appropriate hypotheses.

Two hypotheses to be tested were:

Ho_1: Increased concentrations of air pollution result in no increase in the cost of hospital service per patient admitted and treated.

Ho_2: Increased concentrations of air pollution result in no increase in the rates of admissions to hospitals for treatment of selected illnesses.

The first hypothesis is for c, the second for r.

The Study Area

Allegheny County, Pennsylvania was selected as the study area. There were 49 locations for measuring sulfur dioxide and 21 locations for measuring atmospheric suspended particulates. These locations had been chosen for various reasons as sites of opportunity: for their proximity to emissions sources, for the varied topography, or for "background" measurement. Sulfur dioxide was monitored by the Huey Plate Sulfation Rate Method and reported as SO_2 ppm, the latter being obtained using a regression equation based on simultaneous measurements at representative locations during the study period by both the Huey Plate Method and the Environmental Protection Agency's standard method for SO_2.[9–11] Particulates were measured as the Coefficient of Haze (COH) or Soiling Index, using modified Unico Model 2800 instruments, and converted to $\mu g/m^3$ of total suspended particulates (TSP), using conversion factors determined at the sampling sites. The data reflected, in 1972, neighborhoods that met air quality standards and neighborhoods that did not. The monthly average temperature was below 50° F for six months of 1972, and above 50° F for the remainder of the year.

The County lies in the central portion of the Pittsburgh, Pa SMSA for which census data are available, providing socioeconomic characteristics of the population by census tract.[12] It had 28 hospitals, 26 of which maintained patient records by consistent entries, and two of which maintained equivalent records, providing corresponding information without revealing any individual's identity.[13]

Air Quality Data Base

To assess the impact of ambient concentrations of SO_2 and TSP upon the hospitalization costs to the population of the County, it was necessary to assign particular quantitative values to the air quality of each of the County's 498 census tracts. The values assigned had to be based upon existing measurements at nearby locations. Several available techniques for assigning values were tested. In one attempt, a second-order response surface was fitted to the existing data, but the error at points of measurement was large. Three separate surfaces were then fitted to data from three parts of the County with some success. When the three analyses were combined, however, extreme gradients of concentrations occurred at the boundaries of the areas, making the total analysis unacceptable. An iterative, weighted average analysis technique, similar to procedures used in objective analyses of meteorological data, was finally adopted.[14] Preliminary to application of the technique, the locations of all monitoring stations, and the centroid of all County census tracts were identified by their Universal Transverse Mercater (UTM) coordinates.[14]

In the analysis technique, the contribution of an observed pollutant concentration to the estimated pollutant concentration within a census tract decreases as a function of the distance between the two locations, until the separation exceeds a specified distance, R, called the radius of influence of the observed pollutant concentration. Thus, all pollutant measurements within a radius R of a given location (census tract) are accorded a weight in computing the interpolated pollutant concentration for that location.

The weighted average of sufficiently nearby measured pollutant concentrations is then calculated to give the estimated pollutant concentration for the census tract, $\hat{\phi}_i$

$$\hat{\phi}_i = \frac{\sum_j W_{ij}\phi_j}{\sum_j W_{ij}}$$

where W_{ij} is a weight function dependent upon r_{ij} and R.

There are many possible choices of function form for W_{ij}. The form used by Cressman,[15]

$$W_{ij} = \frac{R^2 - r_{ij}^2}{R^2 + r_{ij}^2}, r_{ij} < R$$

$$W_{ij} = 0, r_{ij} \geq R$$

was chosen based upon successes in previous interpolations of atmospheric variables.

A three-step iterative procedure was adopted to analyze the entire area of Allegheny County. On the first iteration, a large value of R, R(1), was used so that at least three observations were included. On the second iteration, a smaller value, R(2), was used. R(2) depended upon the average density of observation over the analysis area. The results of the first iteration were incorporated into the second analysis by a preassigned weight. Finally, R(3), a very small value, was used with the previous iteration to help regain the observed values within the analysis field.

Letting ν indicate the iteration, and $W_{ij}(\nu)$ the Cressman weight function using $R(\nu)$, the following equation shows calculation of the census tract pollutant level:

$$\hat{\phi}_i(\nu + 1) = \frac{\overline{W}\,\phi_i(\nu) + \sum_{j=1}^{n} W_{ij}(\nu)\,\phi_j}{\overline{W} + \sum_{j=1}^{n} W_{ij}(\nu)}$$

where $\phi_i(0) = 0$, and $\overline{W}$ is a small positive constant.

On the first iteration, this equation estimated a pollutant concentration for each census tract. In the next iteration, if there are no observations within distance R(2) of a census

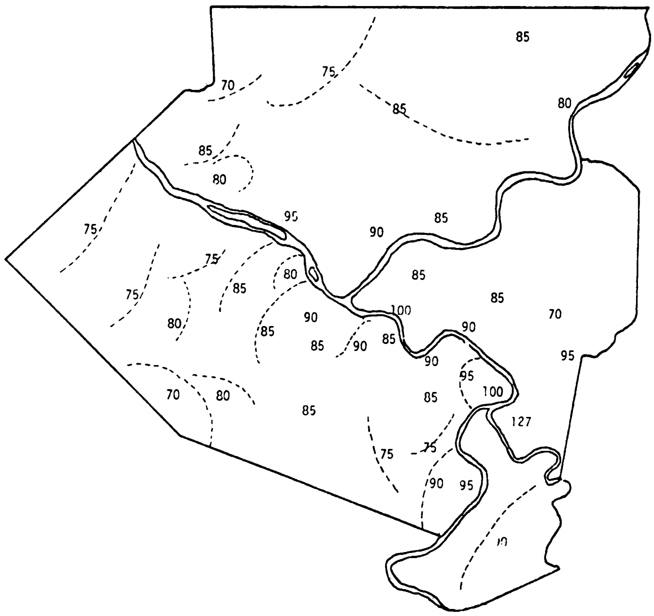

FIGURE 1—SO_2 Levels (1972 yearly average), Allegheny County, Pennsylvania, $\mu g/m^3$

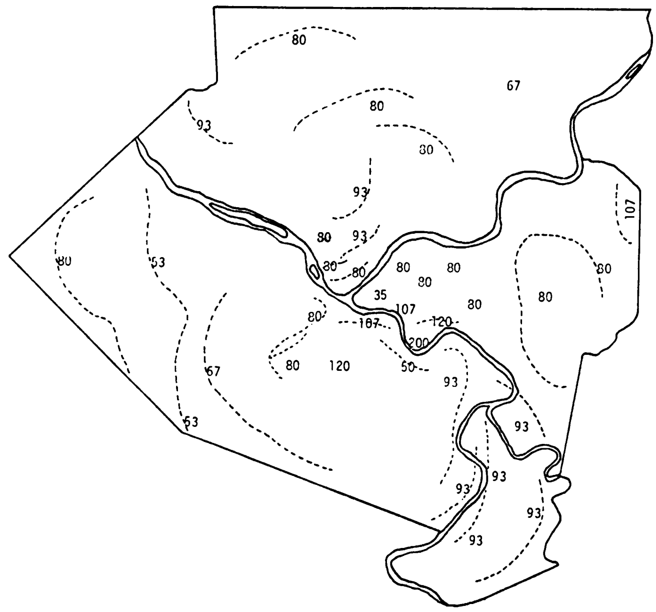

FIGURE 2—Particulates Levels (1972 yearly average), Allegheny County, Pennsylvania, $\mu g/m^3$

tract, the first interpolated value was retained. Otherwise, the new interpolation would depend primarily upon nearby observations but, for continuity, would weakly incorporate the first estimate, using a multipler $\overline{W}$. In this study, $\overline{W}$ = 0.05, gave satisfactory results.

Figure 1 shows approximate contours of levels of SO_2 drawn on a map of Allegheny County, using the estimated levels by census tracts. Figure 2 shows approximate contours of levels of particulates. The effects of uneven terrain and known emission sources are quite evident.

Pollutant concentrations assigned to a few selected census tracts are shown in Table 1. A complete listing of the estimated air quality by census tract is given in Reference 14.

Census tracts were sorted into three SO_2 and three particulate levels to establish subpopulations for comparison. Level 1 tracts met the standards. Level 2 tracts exceeded the standard enough to have a possible effect. Level 3 tracts exceeded the bound of Level 2. Class levels all shown in Table 2.

TABLE 1—1972 Annual Average Concentrations of SO_2 and Particulates at Five Selected Census Tracts in Allegheny County, Pa.

Census Tract	UTM Coordinates: Easting	UTM Coordinates: Northing	SO_2 $\mu g/m^3$	Particulates ($\mu g/m^3$)
101	548.6	4476.7	85.8	87.9
1604	587.9	4474.3	141.0	84.9
2003	579.3	4478.3	40.6	74.9
4980	597.2	4464.4	155.0	88.5
5514	597.8	4466.2	97.9	124.7

Hospitalization Data Base

Three classes of diseases were considered: respiratory diseases, suspect circulatory diseases, and comparison circulatory diseases.* All diseases were identified by the ICDA-8 Code.[16] Respiratory diseases included in the data base were ICDA-8 numbers 462 through 515.9, except 508.1 (polyp of vocal cords or larynx). Suspect circulatory system diseases were numbers 410-414.9; 427-429.9; 435; 435.9; and 436.9.[16] These include the ischemic heart diseases, symptomatic heart diseases, transient cerebral ischemia, and acute cerebrovascular disease. Comparison circulatory system diseases were numbers 390-404; 420-426; 430-434.9; 436; 437-448; 450-458.9; and 580-584. These include rheumatic fever, chronic rheumatic heart diseases, hypertensive diseases, certain cerebrovascular diseases, diseases of arteries, arterioles, and capillaries, and diseases of veins. The last category identified nephritis and nephrosis. There were only a few cases of these diseases, and they were included in the comparison diseases.**

TABLE 2—Class Levels for SO_2 and Particulates

Class Level	SO_2: $\mu g/m^3$	SO_2: ppb	Particulate: $\mu g/m^3$	Particulate: Level
1 (low)	<80	< 30	<76	1 (low)
2 (medium)	80 – 99.3	30 – 37.2	76 – 115	2 (medium)
3 (high	>99.3	>37.2	>115	3 (high)

*Consultation with Edward Haag, MD, Human Effects Research Branch, Environmental Protection Agency.

**A breakdown of all diseases by diagnosis and number of cases is included in Reference 14.

TABLE 3—Hospital Admissions Data Base for Diseases under Study

Disease Class	HUP Hospitals			D_o[c]	Non-HUP Hospitals		
	Total	Analyzed	In County		Total	Analyzed	In County
Respiratory	11,550	9,144	9,144				
Circulatory	21,133	3,000[a]	16,285	2	30	30	30
Comparison	5,049	1,198[b]	3,864	1	56	48	48
Total	37,732	13,342	29,293		86	78	78

[a]Sample sized to limit α and β risks to 0.05, where α = risk of claiming a difference greater than 0 when it is really 0; β = risk of claiming a difference = 0 when its really Do.

[b]Sample sized to limit α risk to 0.01, β risk to 0.05.

[c]Do = Difference in hospitalization rates, admissions per 1,000 population, for subpopulations, the detection of which involves risks α and β.

HUP = Hospital Utilization Project.

Records (with case history but without names) of all admissions for treatment of these diseases in 1972 were obtained from the 28 hospitals in Allegheny County. The total number of records was 37,818. While all respiratory disease cases residing within the County were analyzed, only random samples of the circulatory diseases were taken for analysis (Table 3). This made it possible to make meaningful comparisons of small subpopulations for respiratory diseases which are known to be affected by air pollution. A sample of 3,000 suspect circulatory disease cases was taken to provide good control of possible sampling errors in detecting differences in hospitalization rates as low as two or more per thousand; a sample size of 1,198 comparison circulatory diseases would provide similar control of sampling error in detecting differences of one or more per thousand.[14]

The random samples were drawn from strata defined by race, sex, and age. Sample allocation to strata was proportional to the number of cases in each stratum. Cases so selected were assigned to the proper census tract, thus identifying their pollutant exposure levels. Of the sample of 3,000 suspect circulatory diseases, 2,311 were from Allegheny County; of the 1,198 comparison circulatory diseases, 917 were in this County.

The total data base for analysis of the three classes of diseases included 12,450 cases from Allegheny County. This is believed to be the best data base ever used for the purpose of estimating hospitalization costs associated with air pollution.

Population Data Base

The population base was developed from 1970 Census data.[17] Table 4 shows the race-age-sex distribution of the population. Here the population is further classified according to the pollutant levels in their neighborhoods of resi-

TABLE 4—Allegheny County, Pennsylvania Population Data Base (in thousands of persons)

Stratum			Pollutant Classification*						
Race	Sex	Age (years)	LL	LM	ML	MC	HL	HM	Total
White	Male	1 to 44	45.566	125.707	20.293	179.964	2.015	72.260	445.80
White	Male	45 to 64	16.042	49.282	5.444	69.628	0.920	29.607	170.92
White	Male	65 to 74	3.097	11.011	0.964	18.059	0.258	7.740	41.73
White	Male	75 or Older	2.110	6.231	0.620	10.317	0.140	4.680	24.20
White	Female	1 to 44	46.941	130.084	20.635	189.213	2.013	76.186	465.07
White	Female	45 to 64	16.971	54.675	5.628	81.353	0.973	34.082	193.68
White	Female	65 to 74	3.829	14.621	1.290	26.160	0.372	10.799	57.07
White	Female	75 or Older	3.524	10.248	0.821	16.993	0.245	6.758	38.59
Black & Other	Male	1 to 44	0.941	5.974	0.441	27.647	0.177	12.352	47.53
Black & Other	Male	45 to 64	0.261	1.473	0.084	8.541	0.023	3.241	13.62
Black & Other	Male	65 to 74	0.145	0.477	0.009	2.880	0.001	1.329	4.84
Black & Other	Male	75 or Older	0.112	0.221	0.021	1.304	0.014	0.560	2.23
Black & Other	Female	1 to 44	1.034	6.455	0.378	32.055	0.136	14.640	54.70
Black & Other	Female	45 to 64	0.307	1.616	0.073	10.049	0.044	3.904	15.99
Black & Other	Female	65 to 74	0.125	0.451	0.030	3.069	0.011	1.451	5.14
Black & Other	Female	75 or Older	0.178	0.236	0.015	1.427	0.001	0.772	2.63
TOTAL			141.183	418.852	56.746	679.259	7.343	280.370	1583.75

*The SO_2 level is indicated by the first letter; the particulates level, by the second: L = low, M = medium; H = high; C = combined M + H. There were no LH or HH census tracts. MC = MM + MH.

dence. The County had six different SO_2-particulate classifications. These are identified by letters. For example, there were 72,260 white males, ages 1-44 years, in the high SO_2-medium particulates neighborhoods (HM).

The Census data were the source of additional population characteristics utilized in this study. For each of the 498 census tracts of Allegheny County, the median income, the fraction of the population with income below poverty level, the fraction married (and maintaining a home together), and the fraction employed in heavy occupations were considered. Heavy occupations were Bureau of Census classes: 5 (Craftsmen and Kindred Workers); 6 (Operatives, except Transport); 7 (Transport Equipment Operatives); and 8 (Laborers, except Farm). The remaining classes, including unemployed, retired, and housewives, were considered to be light occupations.

Ideally, both hospitalization and population data should be based on the same year. Since 1972 is only two years from 1970, it has been assumed that population changes and shifts occurring over that short period of time would not materially affect the results of the analysis.

Data Analysis and Results

The statistical analyses of the data were carried out to decide, within controlled error levels, whether subpopulation hospitalization rates and length of stay for various conditions were similar or different. Hospitalization rates were defined as the number of admissions per year per 1,000 population.

The conclusion that any detected differences among subpopulations are directly attributable to air pollution may not, of course, be based solely on statistical tests, since the subpopulations may possess a number of other known or unknown characteristics which were in no way experimentally controlled to allow the difference to be associated solely with the pollutant level. The inability to conclude causation even though association has been proven is a weakness of all observational studies.[2]

Many previous health studies, however, have established that the respiratory diseases under study increase in prevalence or incidence among populations exposed to SO_2 and particulate air contamination. Additional studies are showing the biological impact of such exposure on living tissues. Given these findings, this study is concerned only with estimation of those increases in hospitalization rates for these diseases that may be occurring among exposed populations, and the translation of such increases to cost. For this purpose, it is necessary only to prove association.

Lesser evidence supports the effects on circulatory system diseases and thus this study is itself a source of further evidence. These diseases were sorted into two groups: suspect diseases (i.e., those which medically would be expected to be influenced by exposure to pollution), and comparison diseases (i.e., those which medically would not be expected to be so influenced). Comparison circulatory diseases were included in the analysis to serve as controls, although there was no certainty that they were totally unaffected by exposure to pollution. Several benefits from the use of controls are discussed by MacMahon.[18]

Effect of Air Pollution on Hospitalization Rates

A regression analysis was performed in which the unit of observation was the census tract. Hospitalization rates were merged with air quality data and population data for the analysis. The rates (in admissions per 1,000 population) were then examined for their relation to several factors: the census tract SO_2 level (ppb); the particulates level ($\mu g/m^3$), and their product (interaction); the median income; the census tract fractions: males, white race, married (and living with spouse), income less than or equal to the poverty level (as defined by the Census, with consideration of the family status, etc.), employed in heavy industry, and the fractions in three age classes (45-64 years, 65-74, and 75 or over).

All variables were treated as continuous variables. The analysis imposed the assumption that the smoking habits of the population did not vary significantly with census tract pollutant level, since data for this characteristic of the population were not available.

Of the 498 census tracts in the County, required Census data were complete for 493, and these were used in the analysis. The effects estimated by regression analysis are shown in Table 5. Their significance is also indicated. Effects such as those of age are not unexpected and were estimated primarily to minimize their influence on the effects of pollutants.

Exposure to pollution appears not to have affected the hospitalization rates for comparison circulatory disease but to have definitely increased admission rates for respiratory diseases. The effects for circulatory diseases appear to fall mid-way between those for respiratory diseases and comparison circulatory diseases. The evidence is thus comparatively strong that increased hospitalization rates for the suspect circulatory system diseases under consideration here are related to exposure to the higher pollutant levels.

The regression analysis was helpful in assessing several population characteristics as factors that might have affected hospitalization rates. According to the analysis, the fraction males was not a significant one. The effect of race, represented by the fraction white in the analysis, was significant: whites appear to have lower hospitalization rates for suspect circulatory system diseases, compared to other races. Rates for comparison circulatory diseases do not appear to be race-dependent. Married persons appear to incur lower rates of hospitalization for both respiratory and suspect circulatory system disease, but not for the comparison circulatory diseases. Persons with incomes at or below poverty level showed significantly higher rates for respiratory and comparison circulatory diseases, but not for suspect circulatory diseases. Admission rates appeared to increase with census tract median income only for comparison circulatory diseases. They also increased with percentage employed in heavy industry, except in the case of suspect circulatory diseases.

While the regression analysis serves to show that there are significant effects of air pollution on hospitalization rates, the model is inadequate to provide cost estimates of those

TABLE 5—Effects of Exposure to Air Pollution, and Other Factors, on Hospital Admission Rates Obtained by Regression Analysis.

	Comparison Circulatory System Diseases		Respiratory Diseases		Suspect Circulatory System Diseases	
	Effect	Significance*	Effect	Significance	Effect	Significance
Intercept	−8.7	0.41	−15.4	0.20	−41.8	0.13
SO_2 level, ppb	−0.008	0.98	0.71	0.053	1.34	0.12
Particulates Level, $\mu g/m^3$	0.012	0.92	0.27	0.049	0.58	0.069
SO_2 × Particulates	0.00009	0.98	−0.008	0.054	−0.015	0.12
Fraction males	5.97	0.38	−1.52	0.85	1.86	0.92
Fraction white	−1.39	0.20	−3.47	0.005	5.41	0.06
Fraction married	1.99	0.49	−10.8	0.001	−24.9	0.0013
Fraction ≤ poverty level income	74.1	0.005	92.1	0.002	35.4	0.42
Median income, $/yr	0.001	0.098	0.00002	0.79	0.00006	0.78
Fraction in heavy industry	6.2	0.019	8.6	0.005	6.95	0.32
Fraction age 65–74	1.01	0.915	−11.1	0.30	7.0	0.78
Fraction age 45–64	7.63	0.11	19.7	0.0003	37.4	0.003
Fraction age 75 or more	32.6	0.003	15.6	0.215	21.0	0.47

*The significance is the theoretical probability of obtaining an absolute value of t (t = estimate of effect ÷ standard error of the estimate) as large or larger than that exhibited by the data under the hypothesis of a zero effect.

effects. There are several weaknesses of the regression analysis.

The model contains at least three variables (fraction married, fraction ≤ poverty level income, and fraction in heavy industry), the levels of which are subject to some control by the individual. As shown by Crocker, this situation could lead to biases in the estimates of the effects of air pollution on incidence of hospitalization.[19] In the model shown, the nature of the bias is either to make the effect for comparison diseases too low and the effects of respiratory and suspect circulatory diseases too high, or vice versa.

In either case, the bias cannot be easily measured or corrected. The biased effect so estimated may even be negative instead of positive. (The regression model for length of stay also contains individual-controllable variables: occupation, smoking habits, and payment class. The effects of pollutant exposure may be similarly biased in this model.)

The analysis uses ecological variables defined at the census tract level and related to the total census tract population rather than to particular subgroups experiencing hospitalization. The census tract air quality measures are interpolated for each of the 498 census tracts using data from the 49 air quality monitoring stations within the County. Thus the assumption, inherent in the analysis, that all these independent variables are measured with negligible error is weak. The regression model also assumes all effects except those of air quality are approximately linear for the ranges of values observed, and that there are no interactions. Both assumptions would be difficult to verify. The results do provide insight regarding other potentially explanatory variables. In estimating costs, however, it would be desirable to avoid the assumptions inherent in the regression model.

The sample size for studying comparison circulatory diseases (1,200 cases) is equivalent to an average of fewer than 2.5 cases per census tract. It is not surprising then that the census-tract-level hospitalization rate has a high variance and does not lead to statistically significant estimates of many of the potential effects. The lack of significance in the statistical comparison does not prove, per se, the null hypothesis of no effect. Small sample or highly variable data both weaken the power of the statistical test to recognize even those effects that do exist.

To avoid some of these difficulties, the hospitalization rate data were also analyzed as a proportionate stratified random sample, and the results used for cost estimation. The 15 age-sex-race categories utilized in stratifying the hospital cases prior to sample selection were collapsed to 12 in the analysis by combining the age classes "65 to 74 years" and "75 and older" into a single "65 years and older" category. This was done primarily to eliminate all zeros in the population data base.

Total hospitalizations were then estimated for each exposure level, and converted to annual rates per 1,000 persons. These estimates are shown in Table 6.

Part of the observed differences in hospitalization rates among the six areas might be attributed to differences in age, sex, and race distribution. To correct for these differences, estimates of hospitalization rates for each area were also computed using the total SMSA population as a standardizing distribution. These adjusted rates are also shown in Table 6 along with their standard errors. The final entries show the median pollutant level for both SO_2 and particulates, and the median family income.

Standard estimation procedures for stratified samples were utilized in computing estimates of hospitalizations by pollutant exposure areas.[14] Since the inferences drawn from the analysis are intended to extend beyond the finite list of 1972 hospitalizations in Allegheny County, no finite population correction factors were used in any of the variance estimates. This is not meant to suggest a claim of external

TABLE 6—Estimated Hospitalization Rates, by Level of Population Exposure to Air Pollution

	LL[a]	LM	ML	MC	HL	HM	Total
Subpopulation Exposed (1000's)	141.183	418.852	56.746	679.259	7.343	280.37	1583.75
Total Hospitalizations, 1972							
Respiratory Diseases	548	2216	281	4191	78	1830	9144
	(22)[c]	(42)	(17)	(51)	(9)	(39)	(43)
Suspect Circulatory System Diseases[b]	951.5	4292	331.2	7276.9	119.9	3343.5	16315
	(80)	(155)	(48)	(179)	(29)	(58)	(157.8)
Control Diseases	184.3	1015.	76.7	1896.9	25.2	713.9	3913
	(27)	(58)	(18)	(70)	(10)	(50)	(60)
Hospitalization Rates (per 1000)							
Respiratory Diseases	3.88	5.29	4.95	6.17	10.62	6.53	5.77
	(0.16)	(0.10)	(0.29)	(0.08)	(1.20)	(0.14)	(0.03)
Suspect Circulatory	6.74	10.25	5.84	10.72	16.32	11.93	10.30
System Diseases	(0.67)	(0.37)	(0.84)	(0.26)	(3.93)	(0.50)	(0.10)
Control Diseases	1.31	2.48	1.35	2.79	3.43	2.55	2.47
	(0.19)	(0.14)	(0.32)	(0.10)	(1.37)	(0.18)	(0.02)
Hospitalization Rates (per 1000) (Age-Sex and Race Adjusted)							
Respiratory Diseases	4.20	5.67	5.98	5.95	9.82	6.19	5.77
	(0.20)	(0.12)	(0.42)	(0.07)	(1.16)	(0.13)	(0.03)
Suspect Circulatory	7.24	10.74	7.59	10.25	14.91	11.42	10.30
Systems Diseases	(0.64)	(0.41)	(0.12)	(0.26)	(4.13)	(0.48)	(0.10)
Control Diseases	1.56	2.50	2.28	2.67	3.73	2.38	2.47
	(0.26)	(0.15)	(0.62)	(0.10)	(1.58)	(0.17)	(0.04)
Median SO_2 blevel, $\mu g/m^3$	66	66	83	83.1	104.9	105.1	
Median Particulates Level, $\mu g/m^3$	64.4	84.6	72.7	85.3	73.5	87.3	
Median Family Income, $	7343	7096	7312	7107	7055	7358	

[a]The SO_2 level is indicated by the first letter; the particulates by the second, L = low, M = medium, H = high; C = combined M + H; MC = MM + MH. There were no LH or HH census tracts.

[b]The fractional values result from estimating, using the sample data.

[c]( ) denotes standard deviation of the value entered immediately above.

validity to any particular population defined statewide or nationally. Ignoring the finite correction factors treats the occurrence of hospitalizations as a random process taking place under conditions defined in terms of the existing population characterized by age, race, sex, and residence in a defined air quality area.

The direct analysis of the stratified sample by age-sex-race groups makes no attempt to determine the effect of factors other than air quality.*** The age-sex-race adjusted rates do differ from direct estimates, indicating that a portion of the observed differences in hospitalization rates at different air qualities may be associated with age-sex-race related differences in subpopulations. The direct estimation procedures used, based on the stratified sample design, lead to conclusions with no assumptions made about the functional form of the relationship. In this analysis, hospitalization rates for comparison circulatory diseases were higher in areas of higher pollution, and this was taken into account in estimating the costs.

Potential Biases

Some cases were lost from the sample because hospitals could not locate their records. Most of the hospitals felt that the patients involved were most likely from outside the County. The worst possible bias that these missing patients could introduce is the bias resulting from their having all belonged to a single air quality subpopulation of Allegheny County. Although this was considered quite unlikely, its effect was examined. The increases in hospitalization rates that would result if all unclassified cases were added to cases from a single air quality subpopulation were computed. The results show that, if the 458 unclassified cases out of 11,550 respiratory disease cases were added to the LL subpopulation, that rate would increase from 3.88 to 7.12. If added to the HL subpopuiation, that rate would increase from 10.62 to 72.99, which is unreasonably high.

A second potential bias could have resulted from persons in Allegheny County entering hospitals in neighboring counties, and thus not being included in the data. Hospitals in neighboring counties were contacted concerning the possibility; all indicated that to their knowledge there were very few if any such cases. On the other hand, substantial numbers of persons in neighboring counties did enter Allegheny County hospitals and the data analysis took this into account.

Excess Hospitalization Costs

The total number of excess hospitalizations for each of the three disease conditions was estimated by comparing the hospitalization rates in the area of the county meeting air

***Such age-sex-race estimates are possible and were made in the process of developing direct and standard population estimates of hospitalization rates by air-quality class.[14]

TABLE 7—Estimated Excess Hospitalization Rates, Excess Hospital Days, and Excess Hospital Costs

Disease	Hospitalization Rates[a] Air Pollutants At or Below Standards	Air Pollutants Above Standards	Excess Hospitalizations Per 1000 Persons 1 yr or older	Total Excess Hospitalizations	Total[b] Excess Hospital Days	Total[b] Excess Costs
Respiratory	3.88	5.96	2.08	3,000	28,205	$1,920,769
Diseases	(0.16)	(0.03)	(0.17)	(254)	(2,386)	(162,490)
Suspect Circulatory	6.74	10.65	3.91	5,640	77,274	5,192,823
System Diseases	(0.57)	(0.11)	(0.58)	(838)	(11,478)	(771,309)
Comparison	1.31	2.58	1.27	1,832	28,214	2,011,642
Diseases	(0.19)	(0.03)	(0.20)	(300)	(4,615)	(329,119)
Total			7.26	10,472	133,693	9,125,241
			(0.64)	(926)	(12,599)	(854,190)

[a]Not age-sex-race adjusted
[b]Based on a population of 1,442,570 persons 1 year old and over residing in those areas not meeting prescribed standards for air quality; assumed average lengths of hospital stay equal to 9.4 days for respiratory disease, 13.7 days for suspect circulatory diseases, and 15.4 days for control diseases; and assumed average costs per day of $68.10 for respiratory diseases; $67.20 for suspect circulatory diseases; and $71.30 for control diseases.

quality standards (SO_2 <80 μg/m³ and particulates <76 μg/m³) with the remaining parts of the county. These comparisons are shown in Table 7. The difference in rates is applied to the population residing in areas that did not comply with standards. Estimates of 3,000 5,640, and 1,832 excess hospitalizations for respiratory, suspect circulatory, and comparison cicrulatory diseases associated with non-complying areas are shown. Standard error estimates are given in parenthesis below each estimated number.

The cost of excess hospitalizations was calculated assuming that length of stay and cost per day did not vary with level of pollution.‡ Average lengths of stay and average cost per day, computed from data provided by 26 of the 28 hospitals in the study, were: 9.4 days for respiratory diseases; 13.7 days for suspect circulatory system diseases; and 15.4 days for control diseases. Corresponding average audited costs per day were: $68.10, $67.20, and $71.30. These costs differ because of the slightly different distribution of the different diseases among the hospitals.

Using these length of stay and cost per day estimates, total excess hospital days and total excess costs were computed, and are given in Table 7 along with their standard errors.

Effects of Air Pollution on Length of Stay in Hospitals

For effect on length of stay, the population considered was the population of hospitalized persons, not that of the County as in the analysis of hospitalization incidence. A regression analysis was made in which the unit of observation was the individual patient. Length of stay was examined for its relation to several factors: sex, race, age, smoking habits, occupation, how hospital bill was paid, median income for patient's census tract, and hospital load (per cent occupancy). The patient's exposure to pollutants, SO_2, and particulates was based upon the census tract levels.

Median income and hospital load were treated as continuous variables. The others were treated as class variables; differences between specific levels or between specific levels and the average over all levels were computed and tested for statistical significance.

The occupation effect was computed as one-half the difference in average length of stay between "light" occupations and "heavy" occupations.‡‡

Smoking habits were categorized into four groups: unknown (for those whose habits were not recorded), nonsmoker (including those who had quit), light smokers (one pack or less a day), heavy smokers (more than a pack a day, or recorded as "heavy" smoker). The effects were computed by comparing in sequence, the unknowns, the nonsmokers, and the light smokers with the average over all four classes.

Four age classes were established: 1–44 years; 45–65; 65–74; and >74. Each of the first three classes were compared with the average of all classes. Patients less than one year old were excluded because of their limited exposure to air pollution.

Five payment classes were coded: 1) those using Medicare, Medicaid, or Government Insurance; 2) Blue Cross or Commercial Insurance; 3) Worker's Compensation or UMW Insurance; 4) other types of insurance; and 5) selfpayment. Each of the first four classes was compared with the average of all classes.

Because pollutant levels of SO_2 were so often closely correlated with levels of particulates, four classes of air quality were coded: 1) low SO_2 and low particulates; 2) low SO_2 and medium particulates; 3) high, or median SO_2 and low

‡These assumptions are not entirely correct; the costs of extra length of stay are estimated separately in a later section.

‡‡The rationale for this comparison was that the "light" occupations were less likely to impose additional pollutant burdens on the persons than were the "heavy" occupations.

TABLE 8—Effects of Exposure to Air Pollution and of other Factors, on Length of Hospital Stay

	Comparison Circulatory System Diseases		Respiratory Disease		Suspect Circulatory System Diseases	
	Effect	Significance*	Effect	Significance	Effect	Significance
Intercept	28.01	0.003	6.6	0.0001	11.8	0.0002
Sex 0.5 (males-females)	−0.23	0.70	−0.2	0.015	−0.17	0.39
Race 0.5 (whites-others)	0.21	0.83	−0.12	0.33	−0.22	0.51
Surgery 0.5 (no surgery-surgery)	−3.2	0.0001	−0.31	0.0002	−0.29	0.15
Median Income, \$ Δstay/Δ\$	0.000	0.85	0.00004	0.07	−0.0000	0.99
Occupation 0.5 (light-heavy)	−1.0	0.34	0.05	0.71	0.04	0.92
Smoking:						
Unknown-avg.	0.15	0.89	−0.27	0.09	−0.70	0.04
Nonsmokers-avg.	1.02	0.38	−0.69	0.0001	0.25	0.50
Light smokers-avg.	0.22	0.93	0.39	0.24	0.18	0.81
Pollutant exposure**						
Level 2 - level 1	2.12	0.50	1.02	0.030	0.47	0.11
Level 3 - level 1	2.8	0.84	0.35	0.033	3.46	0.14
Level 4 - level 1	2.3	0.77	1.15	0.38	0.61	0.009
Age (years)						
(1-44) - avg.	−2.4	0.25	−4.29	0.0001	−2.5	0.001
(45-64) - avg.	0.96	0.51	−0.02	0.89	−0.14	0.77
(65-74) - avg.	−0.19	0.90	1.82	0.0001	0.97	0.05
Payment class**						
1 - avg.	2.3	0.54	0.12	0.77	1.25	0.34
2 - avg.	0.01	0.99	−0.43	0.29	1.54	0.20
3 - avg.	−8.7	0.39	0.20	0.90	−0.24	0.94
4 - avg.	8.6	0.32	0.70	0.19	−6.99	0.05
Hospital % of occupancy, Δstay/Δ%	−0.16	0.11	0.05	0.0004	0.01	0.72
Number of Cases	951		8,993		2,312	

*The significance is the theoretical probability of obtaining an absolute value of t (t = estimate of effect ÷ standard error of the estimate) as large or larger than that exhibited by the data under the hypothesis of a zero effect.
**See text for explanation of levels and classes.

particulates; 4) high or medium SO_2, and high or medium particulates. These classes covered all patient exposures. Each of the last three classes was compared with the first, which was air that met standards.

The effect of sex was determined as one-half the differences between males and females (a negative effect thus would indicate a greater length of stay for females). The effect of race was calculated as one-half the difference between whites and other races.

The effects estimated by the regression analysis are shown in Table 8. All effects are listed, and their significance is also indicated. However, effects such as those of age and smoking were not unexpected and were estimated primarily to minimize their influence on the effects of pollutants.

Exposure to air pollution appears not to have affected length of stay for comparison circulatory diseses. The comparisons for such diseases are not significant. The comparisons for respiratory diseases show that length of stay for pollutant levels 2 and 3 are significantly higher than for level 1. The comparisons for suspect circulatory diseases show that length of stay for pollutant level 4 is significantly greater than for level 1, and that lengths of stay for pollutant levels 2 and 3 are enough greater than level 1 to merit further study.

Excess Length of Stay Costs

The additional costs to persons in Allegheny County for 1972 were estimated in two steps: the number of hospitalizations for respiratory diseases and for circulatory diseases at each of the three pollutant levels greater than standards was multiplied by the extra length of stay (Table 8) if significantly greater than zero, to obtain the total extra days of hospitalization (Table 9). The total was 10,744 days in 1972. This was then converted to 1972 dollars by multiplying by the audited average cost per day. The total estimated cost to residents of Allegheny County in 1972 due to increased length of stay in the hospital was thus found to be \$731,697.

Conclusions

The results of this study of the hospitalization costs of exposure to levels of SO_2 and particulates in excess of prescribed standards suggest that, in 1972, subpopulations of Allegheny County, Pennsylvania so exposed incurred significantly greater hospitalization costs. Compared to subpopulations of the county living in clean air neighborhoods, subpopulations living in polluted air neighborhoods incurred

TABLE 9—Estimated Additional Costs of Hospitalization Due to Exposure to Air Pollution

Diseases	Air Pollution Levels: At or Below Standards	Low SO_2 Med Part.	Med & High SO_2 Low Part.	High SO_2 Med & High Part.	Total
Respiratory					
Estimated additional days	0	2260 (865)	0	6924 (2228)	9184 (2390)
Estimated additional cost, $	0	153906 (58960)	0	471524 (151700)	625430 (162755)
Suspect Circulatory					
Estimated additional days	0	0	1560 (686)	0	1560 (686)
Estimated additional cost, $	0	0	106267 (46125)	0	106267 (46125)
Total additional days		2260 (865)	1560 (686)	6924 (2228)	10744 (2331)
Total additional costs, $		153906 (58960)	106267 (46125)	471524 (151700)	731697 (169164)

( ) standard deviation

increased rates of hospitalization and increased length of stay for treatment. For the 1.6 million person in the County, the conservatively estimated cost of increased rates of hospitalization was $9.1 million; the cost of increased length of stay, $0.7 million. These add to a total cost of $9.8 million for the year.

Estimated hospitalization rates were corrected for differences in age, sex, and race distributions among the six different air quality subpopulations of the County. Effects of other factors were assessed. For hospitalization rates, these included the subpopulation median income, fraction below poverty-level income, fraction married, and the fraction employed in heavy industry. For length of stay, the effects of patients' smoking habits, occupation, and type of hospitalization insurance were considered, along with the median income of his or her area of residence and the per cent occupancy of the hospital.

REFERENCES

1. Ridker, RG: Economic Costs of Air Pollution, Praeger, New York, 1967.
2. Lave LB, Seskin EP: Air pollution, climate and home heating, Am J Public Health, 62:909 1972.
3. Park WR: The Economic Impact of SO_2 Emissions in Ohio, Midwest Research Institute, Kansas City, MO.
4. Waddell TE: The Economic Damages of Air Pollution, Environmental Protection Agency, Report No. EPA-600/5-74-012, 1974.
5. Sterling TD, Phair JJ, Pollack SV, et al: Urban morbidity and air pollution, Archives of Environmental Health, August 13, 1966, pp 158-170.
6. Sterling TD, Pollack SV, Phair JJ: Urban hospital morbidity and air pollution, Archives of Environmental Health, September 15, 1967, pp 362-374.
7. Sterling TD, Pollack SV, Weinham J: Measuring the effects of air pollution on urban morbidity, Archives of Environmental Health, April 18, 1969, pp. 485-494.
8. Jaksch JA, Stoevener HH: Outpatient Medical Costs Related to Air Pollution in the Portland, Oregon Area, EPA Office of Research and Development, Report EPA-600/5-74-017, 1974.
9. Huey NA: The Lead Dioxide Estimation of Sulfur Dioxide Pollution, JAPCA, 18(9): 610, 1968.
10. Boulerice M, Brabant W: New PbO_2 support for the measurement of sulfation, JAPCA, 19(6): 432, 1969.
11. Allegheny County Bureau of Air Pollution Control, Pittsburgh, PA.
12. U.S. Bureau of Census, Census of Population and Housing, 1979, Census Tracts, Final Report PHC-(1)-162, Pittsburgh, PA SMSA, 1972.
13. Hospital Utilization Project, 400 Penn Center Blvd., Pittsburgh, PA.
14. Carpenter BH, Chromy JR, LeSourd DA, Bach WD: Health Costs of Air Pollution Damages, Environmental Protection Agency Health Effects Research Laboratory, Research Triangle Park, NC, 27711, Report No. EPA-600/5-77-006, February 1977.
15. Cressman GP: An operational objective analysis system, Mon Wea Rev 87:367-374, 1959.
16. Eighth Revision, International Classification of Diseases, U.S. DHEW, Public Health Service, Publication No. 1693, December 1968.
17. U.S. Bureau of Census, General Social and Economic Characteristics, U.S. Summary, PC(1)-Cl, Appendix B, pp APP-29-32, 1970.
18. MacMahon, B, Pugh TF, Ipsen J: Epidemiologic Methods, Little, Brown and Company, Boston, 1960.
19. Crocker TD: Multidisciplinary Research in Experimental Economics: Two Examples, In Glantz M: (ed.), Multidisciplinary Research Related to the Atmospheric Sciences. National Center for Atmospheric Research, Boulder, Co., Pub. No. NCAR/3141-78/1, June 1978, pp 165-191.

ACKNOWLEDGMENTS

This research was sponsored by the U.S. Environmental Protection Agency (EPA) under Contract No. 68-01-0427. Dr. Donald G. Gillette, Criteria and Special Studies Office, served as Project Officer and provided guidance and direction; Dr. Fred Abel initiated the study and identified several of its methodological concepts; Mr. TE Waddell, of EPA's Office of Planning and Review, served as project coordinator and assisted in development of the data base. The cooperation and help extended by the representatives of the hospitals contacted during the study is gratefully acknowledged.