The aim of this report is to determine whether changes in air pollution have an effect on health. Climate and home heating variables are included to see whether they may be involved. These studies indicate a close association between mortality rates and air pollution and lead to the conclusion that mortality rates could be lowered by abating pollution. Estimates of economic benefits from improved health are discussed.

Air Pollution, Climate, and Home Heating: Their Effects on U.S. Mortality Rates

Lester B. Lave, Ph.D and Eugene P. Seskin, M.S.

Introduction

We have investigated the health effects of air pollution and reported several sets of results (Lave and Seskin, 1970, 1970a, 1970b, 1971, Lave, 1972). The basic approach is to explain differences in the mortality rates among U.S. cities by the level of air pollution and socioeconomic variables. The aim of this work is the estimation of the benefits of pollution abatement.

Two objections were raised to the prior results. One concerned the fact that weather is known to affect health, but no meteorological variables were included in the analysis.[1] The other concerned personal pollution arising from home heating sources. Either of these factors might be the "true" cause of the observed association between air pollution and ill health; if so, abating air pollution would have little effect on health. Neither the literature relating ill health to weather or to home heating equipment is developed sufficiently well to suggest a physiological mechanism associating them to chronic disease. Thus, we are confined to a search for significant, plausible relationships.

The Statistical Model

Our goal is the determination of the effect of changes in air pollution on health. To answer this question, confounding factors must be accounted for or held constant. In explaining variations in the mortality rate across cities, one must hold constant many socioeconomic and other variables. We hypothesize that the mortality rate in a city can be written as in equation (1),

$$(1)\ MR_i = MR(P_i, S_i, C_i, H_i, e_i)$$

where MR_i is a mortality rate in city i, P_i is one or more measures of air pollution in city i, S_i is a vector of measures of socioeconomic status in city i, C_i is a vector of measures of the climate in city i, H_i is a vector of variables representing the home heating characteristics in city i, and e_i is an error term for omitted variables.

To estimate equation (1), we assume that the complex relation can be approximated by a linear function, i.e., the mortality rate is a linear function of air pollution, socioeconomic variables, climate factors, and home heating characteristics. Other functional forms were examined and it was found that the linear form is as good as any (Lave, 1972). A discussion of the general model, along with problems stemming from errors in variables and omitted variables, is given elsewhere (Lave and Seskin, 1970b).

Since there are no data which would allow us to hold the confounding factors constant, we must account for their effects statistically. To do this, we employ multivariate regression analysis. Given the linear specification, and a few plausible assumptions, simple least-squares provides best linear unbiased estimates of the effect of each variable.[2]

Method

In a previous study (Lave and Seskin, 1970a) we determined the "best" set of regressions for a number of mortality rates. The specification was reestimated with data from another year. It was concluded that air pollution had important effects on mortality, even when socioeconomic variables were controlled. In the present analysis we add sets of "heating" variables in order to investigate the importance of the indoor environment on mortality and to examine the interactions of the heating variables with the pollution variables. We also add climatic variables to examine their influence on the observed relationships.

The heating variables are added in sets grouped according to the type of heating equipment, type of heating fuel, type of water heating fuel and a measure of the number of air-conditioned homes. More precisely we added each group and tested the results to see if the explanatory power (R^2) of the regression was increased significantly.[3] We did this first with the heating equipment variables since they were thought to be of primary concern and since they made generally the greatest contribution to explanatory power. If the contribution of this category was statistically significant, we continued adding the remaining sets of variables until there was no longer a significant increase in R^2. When heating equipment did not prove to be a significant factor in the first instance, we tried the other classes of heating variables and followed the same procedure. This constituted the first portion of the analysis for each mortality rate.

In a parallel investigation, we added climatic variables to the "best" 1960 regressions. Only those climatic variables which made a significant contribution to explanatory power were added. Finally, we added heating variables in the same manner as described above to the 1960 regressions with the weather variables present.

This sequential estimation method is intended to bracket the effect of the two additional sets of variables on the air pollution (and socioeconomic) parameter estimates. If either home heating characteristics or climatic factors are the "true cause" of ill health (and air pollution is merely a spurious effect), this estimation procedure is designed to show it. Care must be taken in using these results to estimate the effect of either home heating characteristics or climate on mortality rates.[4]

The Data[5]

We collected data on 117 SMSAs. Air pollution data are reported by the U. S. Public Health Service. Suspended particulates and total sulfates are measured for biweekly periods in micrograms per cubic meter ($\mu g/m^3$). Observations are collected on the biweekly minimum and maximum readings and the annual arithmetic mean.[6] There are a number of difficulties with these data. The measuring instruments change over time and across cities and some instruments have little reliability. In addition, the data are generally for a single point in a vast geographical area. Since pollution concentration varies greatly with the terrain, it is a heroic assumption to regard the figures as representative of an entire SMSA in making comparisons across areas.

Climatological variables are reported by the U. S. Department of Commerce.[7] While there is little difficulty with regard to the actual measurement of most of these variables, the observations are for a single point and may not be characteristic of an entire area.

Mortality data are reported in *Vital Statistics of the United States*. These include the total death rate and a breakdown of the total death rate into age specific categories, including various categories of infant death rates (as a ratio to live births). One problem with the infant death rates is that a classification such as fetal deaths may not be reported uniformly well across all areas.

The "heating" variables are reported in the *Census of Housing*.

Finally, the socioeconomic data are taken from the 1960 census as reported in the *County and City Data Book*.

The variables which we use along with their means and standard deviations are reported in the footnote to Table 1.

An Overview of the Results

In this paper we present a detailed analysis of the total mortality and infant death rates. We have also analyzed disease specific mortality rates in a longer version of the paper (available from the authors). We summarize all results in what follows.

In general neither climate nor home heating variables cause the air pollution variables to lose significance. While there are individual pollution coefficients which do lose significance, the coefficients are quite stable. An exception occurs when home heating fuels are added. These variables are associated closely with measured air pollution and both pollution and heating fuel variables tend to become insignificant. For example, the simple correlation between minimum sulfates and "Coal" fuel is .41. Apparently, the type of fuel used for home heating is a major contributor to the air pollution level in the city. Note that this interpretation does not mean that the previous association between air pollution and mortality is disproved, but rather that it is made more specific by directing the association to home heating fuels, rather than at all sources of air pollution.

The socioeconomic variables are correlated with climate and home heating variables. There is an interaction between the home heating variables and population density and percentage of poor families; the latter two variables have some tendency to lose significance when the heating variables are added.

Climate and home heating variables interact very little in the regressions, although there is some indication that the variables may act as surrogates for each other. Adding the two sets of variables simultaneously changes the results little from adding them sequentially.

For most of the mortality rates, the set of heating equipment variables adds significantly to the explanatory power of the regression. Generally, the types of equipment are associated with decreased mortality rates. When heating fuels are present, they tend to be related positively to the mortality rates. The presence of water heating fuels (and obviously hot water) tends to have a negative effect on the various mortality rates. The air-conditioning variable is seldom important.

Specific Results

The regressions relating to total and infant mortality are reported in Tables 1 and 2 respectively. Regressions 1-1 is written out in equation (2),

$$(2)\ MR = 19.607 + \underset{(2.53)}{.041}\ \text{Mean } P_i + \underset{(3.18)}{.071}\ \text{Min } +$$

$$+\ S_i + \underset{(1.67)}{.001}\ P/M_i^2 + \underset{(5.81)}{.041}\ \%\text{N-W}_i + \underset{(18.94)}{.687}\ \% \geq 65_i + e_i$$

where "Mean P" is the arithmetic mean of the 26 biweekly suspended particulate readings, "Min S" is the smallest of the 26 biweekly sulfate readings, "P/M^2" is the population density in the SMSA, "%N-W" is the percentage of the SMSA population who are nonwhite, "% ≥65" is the percentage of the SMSA population who are 65 and older, and "e" is an error term. This regression explains variations in the total mortality across 117 SMSAs extremely well, since 82.7 per cent of the variation is explained ($R^2 = .827$). Each of the coefficients except population density is extremely significant (as shown by the t statistics in parentheses below the coefficients). As expected, increases in each of the variables would lead to an increase in the total mortality rate.

The percentage of older people is the most important variable in equation (2). A one percentage point increase in the proportion (multiplied by ten, × 10) of people 65 and older (raising the mean from 83.93 to 93.93) is

Table 1—Total Mortality

	1-1	1-2	1-3	1-4	1-5	1-6	1-7	1-8	1-9	1-10	1-11	1-12	1-13	1-14
R^2	.827*	.868	.880	.893	.837	.920	.909	.849	.887	.890	.906	.850	.923	.916
Constant	19.607	21.439	25.901	25.976	6.864	24.086	26.182	44.447	37.362	47.732	41.895	37.975	41.192	45.331
Pollution:														
Mean P	.041	.040	.041	.029	.041	.015	.020	.030	.038	.006	.022	.031	.012	.018
	(2.53)†	(2.65)	(.94)	(2.11)	(2.63)	(1.07)	(1.44)	(1.86)	(2.52)	(.35)	(1.53)	(1.89)	(.80)	(1.24)
Min S	.071	.066	.025	.057	.060	.026	.023	.056	.055	.023	.038	.055	.025	.037
	(3.18)	(3.11)	(1.19)	(3.06)	(2.67)	(1.35)	(1.17)	(2.53)	(2.70)	(1.11)	(2.00)	(2.48)	(1.28)	(1.96)
Socioeconomic:														
P/M^2	.001	-.0001	.001	.0004	.001	.001	.001	.0004	-.0001	.001	.001	.0005	.001	.0004
	(1.67)	(-.20)	(2.75)	(.82)	(1.75)	(1.30)	(1.57)	(.98)	(-.21)	(2.58)	(1.23)	(1.05)	(1.22)	(.87)
% N-W	.041	.038	.045	.031	.048	.040	.038	.054	.054	.053	.044	.055	.047	.046
	(5.81)	(5.26)	(6.32)	(4.48)	(6.49)	(5.08)	(5.57)	(6.03)	(6.60)	(5.99)	(5.35)	(6.03)	(5.08)	(5.67)
% ≥ 65	.687	.610	.646	.619	.676	.602	.612	.664	.632	.641	.629	.663	.609	.621
	(18.94)	(16.96)	(19.48)	(19.34)	(18.95)	(18.18)	(18.89)	(17.42)	(17.79)	(18.67)	(19.54)	(17.32)	(17.70)	(19.25)
Climate:														
Rain								.001	-.001	.001	-.0003	.001	-.0001	-.001
								(1.20)	(-1.89)	(1.08)	(-.50)	(1.27)	(-.12)	(-.75)
H 1AM								-.285	-.181	-.284	-.184	-.272	-.185	-.216
								(-2.75)	(-1.65)	(-2.77)	(-1.92)	(-2.58)	(-1.69)	(-2.13)
Max ≥ 90								-.082	-.102	-.063	-.073	-.069	-.062	-.097
								(-3.81)	(-3.44)	(-2.28)	(-3.51)	(-2.53)	(-1.78)	(-3.42)
Heating Equipment:														
Steam		17.825				4.312	15.024		18.098				2.356	6.342
		(4.78)				(.47)	(3.36)		(4.49)				(.25)	(.70)
Floor		-3.552				-5.394	-5.324		-4.475				-5.162	-6.940
		(-.63)				(-.98)	(-1.00)		(-.77)				(-.90)	(-1.25)
Elec.		11.816				-45.735	-53.774		15.033				-16.909	27.261
		(1.11)				(-.68)	(-.81)		(1.47)				(-.23)	(2.36)
Flue		12.888				9.115	19.307		22.870				12.806	13.638
		(2.37)				(.94)	(3.01)		(3.92)				(1.30)	(1.58)
N Flue		5.792				-.660	6.798		22.040				6.064	9.979
		(1.32)				(-.11)	(1.56)		(3.84)				(.83)	(1.62)
None		-16.663				266.641	279.864		-12.474				180.038	-7.525
		(-.78)				(2.76)	(2.88)		(-.60)				(1.66)	(-.35)
Heating Fuel:														
Oil			1.365			2.978	-7.613			-1.046			2.426	
			(.64)			(.40)	(-2.37)			(-.41)			(.32)	
Coal			23.827			12.028	15.903			23.455			13.665	
			(5.47)			(1.23)	(3.46)			(5.40)			(1.36)	
Elec			2.351			56.049	41.914			2.462			31.835	
			(.32)			(1.12)	(.84)			(.33)			(.59)	
B. Gas			-42.730			-11.813	-65.093			-.266			5.346	
			(-1.63)			(-.21)	(-2.28)			(-.00)			(.10)	
Other			-4.289			-47.982	-8.905			2.693			-40.864	
			(-.18)			(-1.41)	(-.36)			(.11)			(-1.19)	
None			8.552			-308.204	-294.875			-29.936			-220.998	
			(.27)			(-2.78)	(-2.67)			(-.86)			(-1.80)	
Water Heating Fuel:														
Elec				-1.417		-12.347					-1.795		-10.964	-7.784
				(-.57)		(-1.59)					(-.69)		(-1.30)	(-1.75)
Coal				39.003		20.522					37.561		20.085	32.237
				(6.45)		(1.51)					(6.41)		(1.46)	(3.42)
B Gas				-43.140		-50.381					-17.592		-41.079	-8.576
				(-1.95)		(-.94)					(-.79)		(-.76)	(-.34)
Oil				11.851		-1.693					8.033		1.285	1.159
				(3.07)		(-.13)					(1.80)		(.10)	(.12)
Other				118.754		164.095					77.834		140.341	81.184
				(1.68)		(1.85)					(1.14)		(1.58)	(1.15)
None				29.984		29.350					43.634		25.963	30.482
				(2.84)		(1.42)					(4.06)		(1.21)	(2.03)
NA-C					15.771	6.298						5.733	2.740	
					(2.58)	(.90)						(.75)	(.34)	

* The coefficient of determination; value of .827 indicates a multiple correlation coefficient of .910 and that 83 percent of the variation in the death rate is "explained" by the regression.

† The t statistic; a value of 1.66 indicates significance at the .05 level, using a one-tailed test.

Footnote to Table 1 (cont'd)
Variables Used in the Analysis

	Mean	Standard Deviation
Air Pollution		
Suspended Particulates (μg/m³)		
Minimum reading for a biweekly period (1960)	45.47	18.57
Maximum reading for a biweekly period	268.36	132.07
Arithmetic Mean (annual)	118.14	40.94
Total Sulfates (μg/m³x10)		
Minimum reading for a biweekly period	47.24	31.28
Maximum reading for a biweekly period	228.39	124.41
Arithmetic Mean (annual)	99.65	52.88
Mortality		
Total death rate (per 10,000)	91.26	15.33
Infant death rate (per 10,000 live births)		
< 1 year	254.03	36.44
< 28 days	187.29	24.52
Fetal	153.15	34.35
Climate		
Average daily maximum temperature (x10)	654.99	79.79
Average daily minimum temperature (x10)	459.73	75.71
Degree Days	4682.53	1968.54
Total Precipitation (inches x100)	3710.45	1309.10
Relative humidity 1:00a E.S.T.	76.81	8.11
Relative humidity 1:00p E.S.T.	56.96	7.39
Average hourly wind speed (x10)	91.71	19.05
Precipitation .01 inch or more (# of days)	109.89	26.74
Snow, sleet 1.0 inch or more (# of days)	8.21	6.62
Heavy fog (# of days)	27.07	18.97
Maximum temperature 90° and above (# of days)	38.23	39.15
Maximum temperature 32° and below (# of days)	27.18	28.96
Minimum temperature 32° and below (# of days)	94.30	49.57
Minimum temperature 0° and below (# of days)	3.50	7.54
Socioeconomic		
Persons per square mile	756.15	1370.54
% nonwhites in population (x10)	125.06	103.98
% population ≥ 65 (x10)	83.93	21.21
% families with incomes < $3,000 (x10)	180.85	65.53
Heating (% / 100)		
Heating equipment		
Steam or hot water	.20	.22
Warm air furnace	.35	.22
Floor, wall, or pipeless furnace	.13	.13
Built-in electric units	.02	.05
Other means with flue	.18	.12
Other means without flue	.12	.18
None	.01	.03
Heating fuel		
Utility gas	.49	.33
Fuel oil, kerosene, etc.	.31	.29
Coal or coke	.11	.15
Electricity	.02	.07
Bottled, tank, or LP gas	.03	.03
Other fuel	.02	.02
None	.01	.02
Water heating fuel		
Utility gas	.54	.28
Electricity	.22	.22
Coal or coke	.03	.09
Bottled, tank, or LP gas	.04	.02
Fuel oil, kerosene, etc.	.09	.16
None	.08	.06
No Air-conditioning	.84	.12

Table 2—Infant Mortality

	< 1 year		< 28 days		Fetal		
	2-1	2-2	2-3	2-4	2-5	2-6	2-7
R^2	.537	.575	.271	.322	.426	.512	.548
Constant	185.802	228.842	149.428	167.274	93.852	84.566	181.312
Pollution:							
Min P	.365 (2.82)	.340 (2.61)					
Mean P			.083 (1.62)	.066 (1.19)			
Min S			.120 (1.82)	.121 (1.74)			
Mean S					.141 (2.67)	.081 (1.46)	.050 (.94)
Socioeconomic:							
P/M²					.003 (1.61)	.002 (.90)	.003 (1.61)
% N-W	.186 (6.52)	.195 (6.72)	.098 (4.04)	.088 (3.19)	.161 (5.33)	.192 (6.42)	.183 (5.60)
% Poor	.157 (3.38)	.163 (3.43)	.056 (1.45)	.075 (1.85)	.125 (2.49)	.211 (3.24)	.184 (3.35)
Climate:							
Rain		.003 (1.37)		.002 (1.08)			
H 1 AM		-.958 (-2.78)		-.453 (-1.39)			-.764 (-1.51)
H 1 PM							-.804 (-1.65)
Wind		.196 (1.54)		.145 (1.29)			
Rain ≥ .01							.294 (2.35)
Fog				-.164 (-1.34)			-.344 (-2.54)
Max ≥ 90°							-.263 (-2.51)
Heating Equipment:							
Steam						30.029 (1.96)	
Floor						2.243 (.10)	
Elec						-38.701 (-.86)	
Flue						-18.443 (-.63)	
N Flue						-53.699 (-2.34)	
None						104.299 (1.16)	

estimated to raise the total death rate 6.87 per 10,000 (from a mean of 91.26 to 98.13). Increasing nonwhites in the population (× 10) by 1 percentage point (raising the mean from 125.06 to 135.06), is estimated to raise the total death rate by .41 per 10,000. If air pollution worsened and either the minimum sulfate level or mean particulate level rose by 1 microgram per cubic meter (ug/m^3), the total death rate would rise by either .71 or .041, respectively.

A difficulty arises in attempting to estimate the relation when a set of the home heating variables is to be added. The variables are defined as the percentage of all homes in an area heated by a particular method, such as "Steam." Since the sum of all variables within a set is identically 100 per cent, adding all variables would preclude inverting the matrix of cross products and make it impossible to derive estimates of the regression coefficients. A simple solution to the difficulty is to exclude one of the variables. The estimated regression coefficients of the included variables are then interpreted as the difference between the coefficient of the variable and the coefficient of the excluded variable.[8]

This difficulty can be clarified by examining regression 1-2, written out as equation (3),

$$(3)\ MR_i = 21.439 + \underset{(2.65)}{.040}\ \text{Mean P} + \underset{(3.11)}{.066}\ \text{Min S} - \underset{(-.20)}{.0001}\text{P/M}^2 + \underset{(5.26)}{.038}\ \%\text{N-W} + \underset{(16.96)}{.610}\ \% \geqslant 65 + \underset{(4.78)}{17.825}\ \%\ \text{Steam} - \underset{(-.63)}{3.552}\ \%\ \text{Floor} + \underset{(1.11)}{11.816}\ \%\ \text{Elec} + \underset{(2.37)}{12.888}\ \%\ \text{Flue} + \underset{(1.32)}{5.792}\ \%\ \text{N Flue} - \underset{(-.78)}{16.663}\ \%\ \text{None} + e_i$$

where the first five variables are defined as above, "% Steam" is the percentage of homes in the SMSA with steam or hot water heating, "% Floor" is the percentage of housing units with floor, wall or pipeless furnace, "% Elec" is the percentage with built-in electric units, "% Flue" is the percentage heating by other equipment with a flue, "% N Flue" is the percentage heated by other equipment without a flue, and "% None" is the percentage of homes in the SMSA without heating equipment. The category "Warm Air Furnace" is excluded and all heating effects are relative to this category. If equations (2) and (3) are compared, one notes that the magnitude and significance of the pollution and socioeconomic variables are essentially the same, except that population density becomes insignificant. Only two of the heating equipment variables ("% Steam" and "% Flue") are statistically significant. Both increase the total mortality relative to the excluded category "Warm Air Furnace".

Because of the numerical analysis considerations, the variable we chose to eliminate from each set was the largest category. However, for purposes of interpreting the results, it is more relevant to consider the effect of, for example, each type of heating equipment relative to the "None" category, rather than relative to the "Warm Air Furnace" category.[9] For equation (3), this reinterpretation means that all types of heating equipment are associated with higher mortality rates than is no heating equipment. In other words, in areas with few heated homes the mortality rate is lower than in areas where many homes are heated. In particular, the categories steam or hot water, warm air furnace, and other equipment without flue have the greatest effect on increasing the mortality rate.

The remaining regressions in Table 1 will now be discussed. We discuss one mortality rate at a time, with primary attention given to the effect of heating and climate variables on the air pollution coefficients.

Total Mortality—For total mortality we ran a complete set of regressions (1-1 to 1-14) including the addition of each set of heating variables with and without the climatological variables. We also ran two "best" regressions (1-7 and 1-14) where only significant sets of variables were added. In the previous discussion, the results from the first two regressions were examined in some detail. Regression 1-3 reports the addition of the heating fuel variables to the basic regression (1-1); the magnitude and significance of the pollution variables decrease. As noted previously, one explanation for this lies in the fact that home heating fuels can be a major source of air pollution. The socioeconomic variables in this regression are essentially unaltered although population density does gain some importance. Relative to "None" (no heating fuel at all, and presumably no heating equipment), it is found that the presence of any heating fuel (except for the "Coal" category) is estimated to reduce the total mortality rate. "Coal" is associated with considerable amounts of soot and other undesirable pollutants. We caution that the logical complement to the fuels, namely heating equipment, is not included in this regression. Regression 1-4 contains the water heating fuel variables in addition to those variables found in regression 1-1. The pollution and socioeconomic variables are essentially unchanged from regression 1-1, although population density loses significance. The presence of water heating fuel (the obvious complement to hot water) tends to be associated with lower total mortality, although, again, "Coal" fuel as well as "Other" types of fuel have contrary implications. The percentage of homes without air-conditioning is added to the basic regression in 1-5. The addition of this variable has little effect on the pollution and socioeconomic variables. The variable itself is positive and significant suggesting that air-conditioning reduces the mortality rate. Regression 1-6 contains all sets of heating variables and again the pollution variables are dominated in the presence of the heating fuels. With all the sets included, all types of heating equipment are associated with lower total mortality, while all types of heating fuel are associated with higher total mortality. Except for "Other" types of fuel, all water heating fuels are also associated with lower total death rates. Regression 1-7 contains the subsets of heating equipment and heating fuel and the results are similar to regression 1-6.

Regressions 1-8 to 1-14 add meteorological variables. Regression 1-8 was determined by entering climate variables into the basic regression. We hypothesized that the pollution measures would remain significant explanatory variables and that the climatological variables would not be very important. These results are borne out by the regression. Of the fourteen climatological variables only two are statistically significant. The first one, humidity, indicates that damp regions have a lower death rate than dryer regions. The second climatic variable indicates that as the number of really hot days increases (days of more than 90°), the death rate falls. The coefficient of precipitation approaches statistical significance and indicates that regions with greater precipitation have a higher death rate. If one compares regressions 1-8 to 1-13 with the corresponding regressions without the climate variables (regressions 1-1 to 1-6), one notices that the presence of the weather variables has little effect. One exception is regression 1-12 in which the air-conditioning variable loses significance in the presence of the weather variables. This supports our contention that this variable may be acting as a surrogate for the climate variables in a region, and when they are included explicitly, it becomes unimportant. Regression 1-14 contains the heating equipment and water heating fuel groups. The result is as expected from looking at the two sets of variables independently (regressions 1-9 and 1-11).

Infant Deaths—The death rate for infants under one year is examined in regression 2-1. The minimum particulate level is the important pollution variable, while the percentage of nonwhites in the population and the percentage of poor families in the population are the important socioeconomic variables. Fifty-four per cent of the variation in the mortality rate is explained across the SMSAs ($R^2 = .537$). No set of heating variables contributed significantly to the explanatory power of the original regression. In regression 2-2 weather variables were permitted to enter. The only weather variable which was significant was the humidity reading (1 AM). It indicates that the mortality rate is lower in regions which have higher humidity.

In regression 2-3 the mortality rate for infants under 28 days is explained in terms of pollution and socioeconomic variables. In this case, the mean level of particulate pollution and the minimum level of sulfate pollution are the important pollution measures, while the percentage of nonwhite and the percentage of poor are the relevant socioeconomic variables. Only 27 per cent of the variation across SMSAs is explained. Again, no set of heating variables contributed significantly to the regression. In addition, no climatological variable was statistically significant for the under 28 day category (regression 2-4). The only effects of the weather variables are to decrease the significance of the mean level of particulate pollution and to increase the significance of the "Poor" variable.

Regression 2-5 explains the fetal death rate in terms of mean sulfates, population density, nonwhites and poor families. Forty-three per cent of the variation across the SMSAs is explained. The addition of the heating equipment variables in regression 2-6 increases the explanatory power of the basic regression significantly (R^2 rises from .426 to .512). These variables tend to decrease the importance of the pollution variables while leaving the socioeconomic variables unaffected. The heating equipment variables indicate that the presence of any type is associated with lower fetal death rates. The effect of the climate variables on the fetal death rate is seen in regression 2-7 and is more profound than for the other infant categories. The pollution variable (mean sulfates) loses significance, while four weather variables appear to be related significantly. Apparently the fetal death rate is lowered by humidity, fog and extreme heat, and is raised by heavy rains.

Summary and Conclusions

To test whether previously estimated relations between U.S. mortality rates and air pollution were spurious, we added variables for home heating characteristics and the climatology of a region. The objective was to determine whether these new variables would cause the estimated effect of air pollution to fall and become statistically insignificant. In general, the air pollution variables were quite stable; there were a few instances when the variables lost significance and a few instances when the new variables increased the significance of air pollution.

In addition to this investigation involving the addition of climatological variables and home heating variables, the effect of air pollution on U. S. mortality has been corroborated by different functional forms, by data from another year, and by an investigation of age, sex, and race specific death rates. In general, the significance of the pollution variables is enhanced by disaggregating the mortality rates.

A rather consistent result which occurred in this investigation was that home heating fuels were added to the regression, air pollution variables tended to lose significance. Elsewhere there is also a preliminary result that when occupation variables were added, some of the air pollution variables lose significance (Lave and Seskin, 1971). These two results do not contradict the association between air pollution and mortality, but rather tend to isolate the nature of the problem. Apparently, home heating fuels can be a major source of air pollution; apparently some occupations are closely associated with the level of air pollution. This explanation is plausible if one notes that the air pollution readings are for one site in an SMSA and are taken from 26 biweekly readings. Other investigators have relied on fuel consumption as a measure of air pollution when good measures of pollution were not available.

These studies make it apparent that there is a close association between mortality rates and air pollution. This investigation strengthens the conclusions cited in a previous work that mortality rates could be lowered substantially by abating air pollution. For example, lowering the measured levels of minimum sulfate readings and mean particulate readings by 10 per cent is estimated to lead to a .897 per cent decrease in the total death rate. A 50 per cent abatement would lower the death rate by 4.485 per cent. Assuming that those who are saved have the same life expectancy of others in their cohort, a 50 per cent abatement in air pollution (specifically in minimum sulfates, minimum particulates, and mean particulates) would result in an increase in life expectancy of about one year for a newborn. As estimated elsewhere, such an abatement would reduce the economic cost of morbidity and mortality by just under 5 per cent (Lave and Seskin, 1970). Thus, such an abatement is probably the single most effective way of improving the health of middle-class families. Note that this middle-class family could do something about smoking, but is powerless to lower its exposure to air pollution (except by leaving the city). The importance of this improvement in health can be assessed by noting that eradicating all cancer would result in lowering the economic cost of morbidity and mortality by 5.7 per cent (see Lave and Seskin, 1970).

Even so, there are many reasons to believe that these estimates are gross understatements of the health cost of air pollution. Chronic diseases generally involve long periods of illness. The economic costs, calculated as the sum of lost work and medical expenditures, grossly understate the amount that would be paid to achieve good health for such a chronically ill period. In addition, death may not result from the chronic illness itself, but rather from one or another complication. For example, chronic bronchitis or emphysema is likely to result in death due to heart disease or pneumonia, rather than from the chronic disease.

Perhaps the only good way to estimate the health costs of air pollution would be to analyze morbidity, rather than mortality data. It seems certain that such an investigation would give a higher health cost since no one can die of emphysema or other chronic illnesses who has not suffered them, but some of the people with chronic illnesses die from other causes. In addition, such an investigation would pick up increases in morbidity rates, such as simple respiratory diseases, which may occur long before death is a possibility. Other, much less severe, illness is known to result from air pollution, but to be unrelated to mortality. For example, eye irritation is a common reaction to acute pol. ution; such costs will never be reflected in mortality statistics (except possibly for accidents).

We have concentrated on health effects of air pollution, without alluding to costs associated with cleaning and deterioration of inert materials, with vegetation and animal damage, and the aesthetic effects associated with living in a dirty, uncomfortable, overcast world. These costs can be substantial, as noted in Lave (1972).

In view of the resources devoted to attaining better health, it seems clear that social welfare would rise by spending the resources to abate air pollution substantially. It is time that we pressed forward with a program of abatement.

References

1. See for example Berke and Wilson (1951) or Climate and Man, 1941 Yearbook of Agriculture (1941).
2. That the least-squares method provides the best linear unbiased estimates is the conclusion of the Gauss-Markov theorem, for which $E(ee') = \sigma^2 I$ and $E(e) = 0$ are the basic assumptions. These assumptions are that the basic model must be linear and that the distribution of the errors must have an expected value of zero and have equal, finite variance over the various observations.
3. Adding additional explanatory variables to a regression cannot lower the explanatory power (R^2) and generally increases it. To test whether the increase in R^2 is statistically significant, the decrease in the residual sum of squares (divided by the number of additional variables) is divided by the residual sum of squares (divided by the degrees of freedom). This ratio of mean reduction in residual squares to mean residual squares is distributed according to F.
4. Since only significant variables were added, and since the resulting parameter estimates were not checked by reestimating the relations with data from another year, one should be hesitant to place a great deal of faith in the parameter estimates for these variables. However, the reported coefficients can be used to gain some notion of the possible range of the estimated effects of these variables. For some of the heating and climate variables, the parameter estimates exhibit little change as the specification is altered. This allows increased confidence that these parameter estimates are unlikely to vary greatly if the relations were reestimated with new data. For other variables, the parameter estimates fluctuate to a greater degree and one can have little confidence in the exact estimates.
5. A number of qualifications and problems are alluded to in this report. These are explored in more detail in Lave (1971).
6. For a few areas it was necessary to use an adjoining year, i.e., 1959, because of incomplete data.
7. For some cities climatological data were not available and we utilized information on the closest area for which we had figures.
8. Suppose $Y = b_o + b_1 s_1 + b_2 s_2 + b_3 s_3$, where $s_1 + s_2 + s_3 = 1$. $S = (s_1, s_2, s_3)$ would correspond to a set of heating variables. Then $S'S$ will be singular and estimation by simple, least-squares techniques, will be impossible. However, substitution is possible: $s_3 = 1 - s_1 - s_2$ and so $Y = b_o + b_1 s_1 + b_2 s_2 + b_3(1 - s_1 - s_2)$; which simplifies to $Y = (b_o - b_3) + (b_1 - b_3)s_1 + (b_2 - b_3)s_2$. Thus, one would estimate $Y = a_o + a_1 s_1 + a_2 s_2$.
9. It is easy to interpret the coefficients relative to any excluded variable. As shown in the previous footnote the estimated coefficients take the form: $a_i = b_i - b_3$, when s_3 is the excluded variable. If we would rather exclude s_2, the relation need not be reestimated, but rather the new parameter estimates can be derived for estimating $Y = c_o + c_1 s_1 + c_3 s_3$, where $c_i = b_i - b_2 = a_i - a_2$ since $a_i - a_2 = (b_i - b_3) - (b_2 - b_3) = b_i - b_2$. Significance tests are a more difficult problem. The value of the t statistic given in the table is used to test the difference between the effect of variable i and the excluded variable; a significant value of t indicates that the two variables have significantly different effects. In the above notation, testing the significance of a_1 is equivalent to testing whether b_1 is significantly different from b_3. If one chose to exclude an s_i whose b_i was quite different from the other b's, all of the a_i would be "significant." In transforming the estimated coefficients to exclude a different variable, the standard error of the coefficient will change and it is not simple to derive the new standard error of the coefficient. For our purposes, it is not as important to derive significance tests for individual coefficients, as it is to determine whether a particular set of variables makes a significant contribution to the explanatory power of the regression; this is done via an F test and the result is unaffected by which variable we decide to exclude.

Bibliography

Analysis of Suspended Particulates, 1957-61. U. S. Public Health Serv. Publ. No. 978 (1962).

Berke, J. and Wilson, V. Watch Out for the Weather. New York: Viking, 1951.

Census of Housing. U.S. Dep. Commerce Publ. 1, Pts. 2-8 (1963).

Climate and Man, 1941 Yearbook of Agriculture. U.S. Dep. Agric. Publ. (1941).

Climatological Data, National Summary. U.S. Weather Bureau 11 (1960).

County and City Data Book. U.S. Dep. Commerce Publ. (1962).

Lave, L. Air Pollution Damage. In: Environmental Quality Analysis, A. Kneese and B. Bower, Eds. Baltimore: Johns Hopkins Press, 1972.

Lave, L. and Seskin, E. Air Pollution and Human Health. Science, 169,723 (1970).

Lave, L. and Seskin, E. An Analysis of the Association Between U.S. Mortality and Air Pollution, working paper, (1970a).

Lave, L. and Seskin, E. Does Air Pollution Shorten Lives? Proceedings of the Second Research Conference of Inter-University Committee on Urban Economics (1970b).

Lave, L. and Seskin, E. Health and Air Pollution: The Effect of Occupation Mix. Swedish Journal of Economics (Mar., 1971).

Vital Statistics of the United States (1960). DHEW Publ. (1963).

Dr. Lave is Head and Mr. Seskin is Research Associate, Department of Economics, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, Pa. 15213.

This research was supported by a grant from Resources for the Future, Inc. and by Fellowship AP48992, Air Pollution Control Office, Environmental Protection Agency. The views and any error are those of the authors and do not necessarily reflect the sponsoring agencies.