Title

THE RELATIONSHIP BETWEEN CANCER INCIDENCE AND TWO POLLUTANTS (TOTAL SUSPENDED PARTICULATE AND CARBON MONOXIDE) FOR THE SAN FRANCISCO BAY AREA

Author

Selvin, S.

Publication Date

1980-06-01

LBL-10847 c2
UC-11

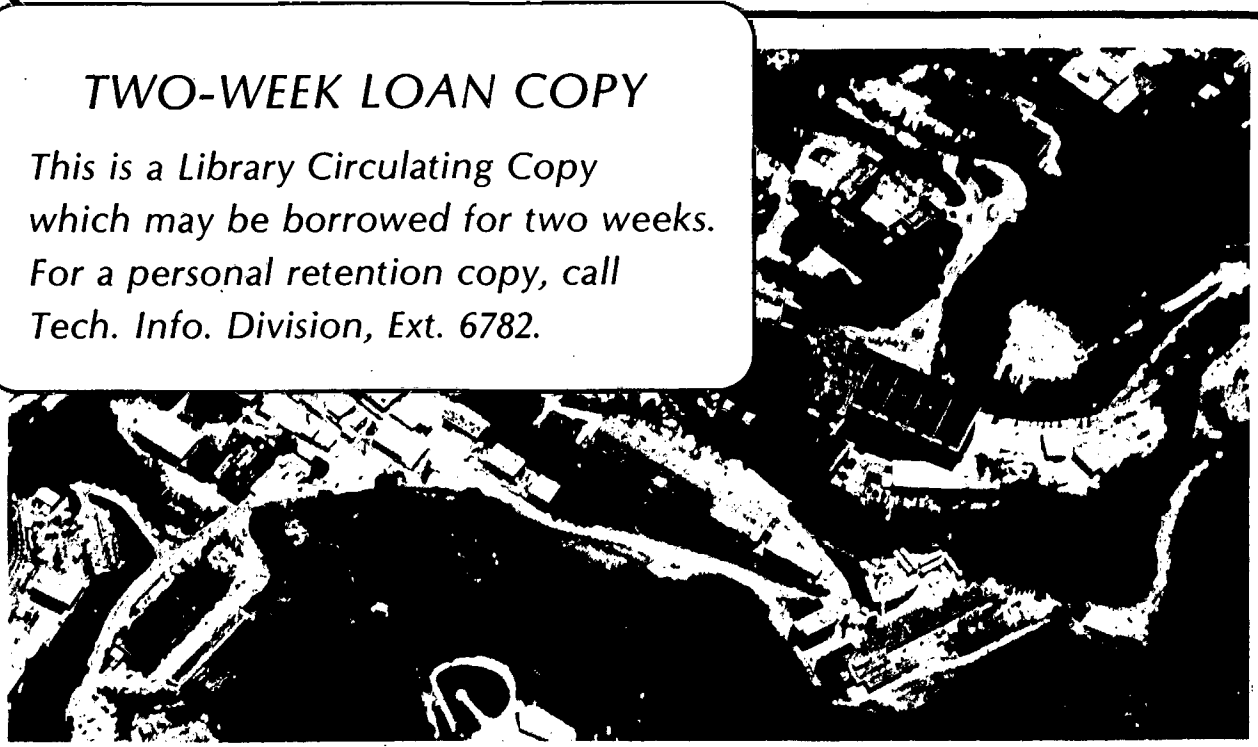
Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

Physics, Computer Science & Mathematics Division

THE RELATIONSHIP BETWEEN CANCER INCIDENCE AND TWO POLLUTANTS (TOTAL SUSPENDED PARTICULATE AND CARBON MONOXIDE) FOR THE SAN FRANCISCO BAY AREA

S. Selvin, S.T. Sacks, D.W. Merrill and W. Winkelstein

June 1980



Prepared for the U.S. Department of Energy under Contract W-7405-ENG-48

DISCLAIMER

LBL - 10847

THE RELATIONSHIP BETWEEN CANCER INCIDENCE AND TWO POLLUTANTS (TOTAL SUSPENDED PARTICULATE AND CARBON MONOXIDE) FOR THE SAN FRANCISCO BAY AREA

S.Selvin, S.T.Sacks, D.W.Merrill and W.Winkelstein

Lawrence Berkeley Laboratory
and
School of Public Health

University of California
Berkeley, California

June 5, 1980

ABSTRACT

Data from the Third National Cancer Survey have been analyzed in conjunction with air quality data, in an investigation of possible associations between site-specific cancer incidence and levels of two pollutants, total suspended particulate (TSP) and carbon monoxide (CO), in the San Francisco Bay Area. Median family income of the census tract of residence was used to stratify the cancer incidence cases according to socio-economic status. The results are consistent with earlier investigations reporting associations between TSP levels and cancer. No such association is found for CO.

The work described in this report was funded by the California Air Resources Board (ARB) under Contract No. A7-185-30, and by the Office of Health and Environmental Research, Assistant Secretary for Environment of the U.S. Department of Energy under Contract No. W-7405-ENG-48.

INTRODUCTION

The data and methods used to study the influence of air pollution on the occurrence of disease have not yet led to widely accepted conclusions. Perhaps the strongest case for an etiologic connection between concentrations of pollution and disease can be made for chronic respiratory disease and stomach cancer.[1-3] Relationships between air pollution and cancer mortality have been noted by several investigators.[3-6] However, the difficulties in accurately estimating pollution exposure, low frequencies of most site-specific cancers, and the lack of data and methods for handling confounding factors make rigorous conclusions concerning general mortality or specific cancer mortality difficult. An alternative method for reaching useful conclusions is the assessment of a series of epidemiologic studies for consistency of results. Such epidemiological investigations typically employ a variety of data sources and statistical methods providing a series of perspectives to help outline and define the relationships, if any, between air pollutants and disease risk. The existing evidence for a relationship between air pollution and cancer occurrence has been reviewed by Higgins.[7]

The following study is another attempt to provide insight into the relationship between pollution exposure and cancer frequency. Like its predecessors, it is not methodologically perfect. However, in an area where imperfect study design is the rule, the present approach possesses some strengths. Relatively small areas (census tracts) within a single geographic entity (San Francisco-Oakland Standard Metropolitan Statistical Area) serve as the fundamental geographic units of study. Age-specific cancer incidence rates for each of 38 individual cancer sites, rather than cancer mortality data, are employed. Air pollution measurements are carefully defined and interpolated to provide estimates of exposure. The statistical methods do not depend on mathematical models (particularly regression models) to adjust for the influences of confounding factors and are purposely chosen

to be simple and straightforward.

MATERIALS AND METHODS

Three types of data were combined to investigate the possible influences of air pollution on the incidence of 38 selected cancers; rates that reflect age-, race- and sex-specific cancer incidence, a measure of socio-economic status, and estimates of exposure to air pollution.

Cancer incidence cases were extracted from the Third National Cancer Survey (TNCS). White male and female residents of the San Francisco-Oakland Standard Metropolitan Statistical Area (SMSA) over age 34 were selected from the 1974 (version 8) master extract. Age, sex, census tract of residence, and cancer site constituted the data record for each individual. A list of the 38 cancer sites selected for this investigation is given in Table 1. Excluding carcinoma in situ and non-melanoma skin cancers, the sample consisted of 31,675 individuals. The inclusion criteria led to slightly fewer cases than reported in the TNCS monograph[8] for some sites. For example, the number of male cases of cancer of the lung, bronchus and trachea is lower by 46 cases (2641 males versus 2687).[8]

Air quality data were obtained from the Environmental Protection Agency (EPA) Storage and Retrieval of Aerometric Data (SAROAD) data bank in the form of yearly summaries (1974-1976). Values of total suspended particulate (TSP) and carbon monoxide (CO) concentration were assigned to each monitoring station. Other pollutants retrieved from SAROAD were not used in this analysis for the following reasons: insufficient monitoring station data (sulfates, total hydrocarbons, non-methane hydrocarbons, ozone and oxidants); and different laboratory methods of measuring the pollutant, producing inconsistent results (sulfur dioxide and nitrogen dioxide). The station values are geometric means of three years of daily TSP and hourly CO measurements, expressed in micrograms per cubic meter. TSP was selected because of the large number of studies employing particulates as a general index of air quality. The CO measurements provide another

"criterion" pollutant with a contrasting geographic pattern (correlation of TSP and CO levels = -.48) and a fairly wide range of levels over the San Francisco-Oakland SMSA. There is also no reason to believe that CO would be associated with cancer occurrence and it, therefore. is viewed as a "control" variable.

The locations of some monitoring stations given in the EPA monitoring station directory were found to be erroneous and were corrected. Then air quality levels for each of the 737 Bay Area census tracts were estimated from the levels observed at individual stations These estimates are weighted geometric means of data from all stations within 100 kilometers of the census tract. The weight of station i is taken to be

$$w_i = n_i \, e^{-\frac{1}{2}(d_i/d_0)^2}$$

where n_i is the number of observations at station i, d_i is the distance from monitoring station i to the tract centroid, and d_0 is a scaling parameter of the order of 5 to 20 kilometers. The geometric mean level for the census tract is estimated as

$$\exp \frac{\Sigma \, w_i \ln x_i}{\Sigma \, w_i}$$

where x_i is the geometric mean of observations at station i. The estimates are based on 39 stations measuring TSP and 21 stations measuring CO. Stations further than five times d_0, which have a negligible weight w_i, were ignored Furthermore, the resulting tract estimate was suppressed for any tract having no stations within a distance equal to three times d_0.

The locations of 18 TSP and 10 CO stations within the San Francisco-Oakland SMSA are shown in Figures 1 and 2. The geometric means and a measurement of the monitoring activity of these stations are given in Figures 3 and 4.

To choose the "optimum" value of d_0 and investigate the accuracy of these weighted averages, a predicted value for each station was generated from observations at the other stations, and compared to the actual value observed at the selected station. This "leave-one-out" method yields a predicted value for each of the Bay Area monitoring stations. The correlation between the predicted and observed values provides an assessment of predictability employing weighted averages derived from monitoring station data. The relationship between the scale parameter d_0 and the correlation coefficient is shown in Figure 5 for both CO and TSP. Predictability (correlation) increases with d_0 to a maximum in the neighborhood of d_0 = 10 km, and then declines. The same pattern is observed for both pollutants. The maximum correlation is 0.82 for TSP and 0.81 for CO.

To demonstrate the importance of selecting the optimum value of d_0, Figures 6 (TSP) and 7 (CO) show the geographic distribution of the interpolated values of TSP and CO for three specific values of the scale parameter d_0 (5, 10 and 20 km). When d_0 = 5 km, the estimated census tract values are dominated by the nearest monitoring station(s). A number of tracts have no station within 3 x d_0 = 15km, and hence no estimate. When d_0 = 20 km, the station values lose their individual identity and the variation across most of the Bay Area almost disappears. The choice d_0 = 10 km generates estimated census tract air quality levels that vary smoothly, and yields nearly the maximum correlation between predicted and observed values when tested on station level data. With the value d_0 = 10 km, ten Bay Area census tracts were more than 3 x d_0 = 30 km from any monitoring station (for either TSP or CO) and were eliminated from the analysis.

Figures 8 (TSP) and 9 (CO) show in greater detail the geographic distribution of estimated pollution levels (with d_0 = 10 km) for the Bay Area census tracts. These values constitute the measure of exposure to air pollutants TSP and CO for this investigation. A brief statistical description of

these estimates is given in Table 2.

The 1969 median family income associated with each census tract, as reported by the 1970 U.S. Census, serves as an indicator of socio-economic status. That is, the general standard of living of each person in the Bay Area is assumed to be reflected by the median family income of that person's tract of residence. The geographic pattern of median family income in the San Francisco Bay Area is shown in Figure 10.

To investigate possible associations between air quality and cancer incidence, the Bay Area census tracts were classified into six categories with respect to income level and exposure to TSP and CO. Three levels of median family income (<$9,700, $9,700 to $12,000, and >$12,000) and two levels of air pollution exposure (above and below the median) were chosen. This choice of income levels produces approximately equal numbers of tracts in each group. While it would have been desirable to use an air pollution classification of at least three levels, the narrow range of variation (see Table 2) precluded more than a two-level classification scheme. Each census tract belongs uniquely to one of these income-exposure groups. The tracts belonging to the low TSP exposure group are not necessarily the same tracts as in the low CO exposure group. The two-way classification of each census tract is depicted in Figures 11 (TSP) and 12 (CO) with respect to the six possible income-exposure categories.

The cancer cases included in the TNCS can be grouped into the six income-exposure categories, since the census tract of residence is recorded for each individual case. Furthermore, these incidence cases can be combined with age- and sex-specific population counts from the 1970 U.S. Census to calculate average annual age- and sex-specific incidence rates for the six income-exposure categories (see Table 3, for example). That is, cancer incidence rates for areas with roughly equal income levels are computed for census tracts experiencing high and low levels of exposure to TSP and CO pollution. This comparison applies equally to all

the cancer sites given in Table 1.

If increased cancer incidence is associated with poor air quality, then within each age-income category the higher rate of cancer should be more or less consistently observed in the group having exposure to higher than median levels of pollution. A simple non-parametric sign test is employed to assess the likelihood that this type of association can occur by chance variation. Under the hypothesis that cancer incidence and pollution are unrelated, the probability that the higher rate within an income group will be associated with the higher exposure category is 1/2. There are 12 comparisons for most cancer sites and sex (one comparison within each of 3 income levels for 4 age-specific groups). The probability, by chance alone, of observing the higher cancer rate in the higher exposure category for all 12 age-income combinations, is $(1/2)^{12} = 0.0002$. Other similar significance probabilities ("p-values") are calculated from the binomial distribution.

As there are 33 male and 35 female cancer sites, the total number of possible comparisons for each pollutant is 12 x 33 = 396 for males and 12 x 35 = 420 for females. Some of the age-sex-income-site categories have no observed incidence cases (for either pollution exposure category), so the actual number of comparisons for each pollutant is 367 for males and 375 for females.

RESULTS

Table 3 gives the incidence rates of cancer of the lung, bronchus and trachea among white males for six income-exposure (TSP) categories for the four age-specific groups. The number of persons "at risk" is also shown. These lung cancer data illustrate the form of analysis applied to all 68 cancer sites (33 male, 35 female) and both pollutants (136 separate analyses in all). In Table 3, nine of the twelve comparisons show the higher rate of lung cancer to be associated with the higher TSP exposure category. The probability of occurrence of this result or a stronger association (significance probability or "p-value") when TSP pollution and lung cancer incidence are unrelated is 0.073. In other words, chance variation is not a very likely explanation for this particular association.

Tables 4 and 5 present summaries of the identical analyses for all 38 specific cancer sites, for both sexes and both pollutants. The degree of association is described in terms of the number of times an increased rate of cancer incidence is found for individuals exposed to the higher (above the median) level of pollution, and in terms of "p-values".

In Figures 13 (TSP) and 14 (CO) the associations between pollution levels and cancer incidence rates are presented in terms of equivalent standard normal deviations (SND). Standard deviations and "p-values" for all sites combined were calculated after first adding together the number of comparisons and the number of positive associations in each case. (To relate accurately the standard normal deviations to the "p-values" the continuity correction factor was employed.) The results are as follows:

Pollutant	Sex	Pos.Assns./Comparisons	p-value	SND
TSP	M	197 / 367	.087	1.41
TSP	F	199 / 375	.128	1.19
TSP	M+F	396 / 742	.036	1.84
CO	M	175 / 367	.826	-.89
CO	F	181 / 375	.765	-.67
CO	M+F	356 / 742	.872	-1.10

There is a 3.6% probability that chance alone could have produced the overall association observed with TSP.

For individual sites the strongest observed associations are as follows:

Poll.	Sex	Site	Pos.Ass./Comp.	p-value	SND
TSP	M	pancreas	11 / 12	.003	2.95
TSP	F	ovary	10 / 12	.019	2.28
CO	M	chron.lymp.leukemia	8 / 9	.020	2.34
CO	F	bones & joints	6 / 7	.063	1.86
TSP	M	lung,bronchus,trachea	9 / 12	.073	1.68
TSP	F	stomach	9 / 12	.073	1.68
TSP	M	testes	9 / 12	.073	1.68
CO	F	Hodgkin's disease	9 / 12	.073	1.68

The associations noted for individual sites should not be overemphasized. An important and often neglected statistical issue arises. When many comparisons are made within the context of a single data set, chance associations with small probabilities of occurrence accumulate to increase the overall probability of observing some differences among a large number of comparisons.

For this data set a reasonable estimate of the overall error is made by assuming each test to be statistically independent. Although a small amount of non-independence exists among the 136 tests employed for each pollutant, the assumption of independence will not be misleading. Only one association (TSP, males, pancreas) has a "p-value" less than 0.01. If all 136 tests with "p-values" of less than 0.01 are declared as significant (i.e. due to non-random influences), the probability of one or more type I errors (for either pollutant) among all tests is about 0.75. (Type I error: the association is declared as real when in fact it arises from chance variation.) Therefore, even at a conservative level of significance (0.01) a large probability exists that one or more of the associations noted in Tables 4 and 5 results from chance variation.

DISCUSSION

Past epidemiological studies of cancer and air pollution have employed mortality data derived from death certificates. Cancer incidence rates, when based on carefully collected data, are generally more effective indicators of cancer risk, particularly for sites with low or rapidly changing case fatality ratios. Employing race-, sex-, age-, and site-specific incidence rates decreases variability due to these four important cancer-related variables and increases the specificity of the analysis, but causes a reduction in the number of cases relevant to each comparison. The trade-off between specificity and sample size always exists and in numerous instances has forced investigators to analyze rather broad age and site categories of cancer in order to include significant numbers of deaths. The 38 cancer sites in this analysis have, for the most part, sufficient numbers of cases to produce rates that reliably reflect cancer risk for the 12 age-income-exposure groups. For example, in the analysis of acute lymphocytic leukemia among females, the rarest observed, 10 cases were employed. At the other extreme, female breast cancer involved 3755 cases.

The accuracy of the rates computed for the 38 cancer sites depends on the completeness of reporting and the precision of diagnosis.[8] Extensive efforts were made to produce the best possible data for the TNCS.[8]

As mentioned, the associations observed for individual cancer sites should not be overemphasized. Neverthess, of the five most significant cancer-TSP associations observed, two (male lung and female stomach) are consistent with previous observations.[3,9] Among the eight gastro-intestinal sites (TNCS codes 500 through 579 in Table 1), we find (for both sexes combined) seven positive associations and one null association with TSP. No such pattern is observed for CO.

The use of CO as a control (since CO is not thought to be

associated with cancer) seems successful. The observed associations between CO and cancer are not more than expected by chance alone. The fact that CO shows no remarkable associations with cancer occurrence for any of the 38 sites gives some confidence that the approach and methods used here do not produce spurious positive results.

The lack of data and/or statistical adjustment for variables that confound possible influences of pollution is a fundamental problem in linking pollution with disease.[10,11] The most common interfering (uncontrolled) variable is cigarette smoking, which certainly influences the occurrence of cancer of a number of sites. If populations exposed to high levels of pollution also have elevated rates of cigarette smoking, a positive association would be observed which is, at least partially, induced by differences in smoking frequency. On the other hand, the lack of any observed association is less easily attributable to the influences of differing smoking rates. No association results when smoking-related increases in cancer rates occuring in populations experiencing low pollution are balanced by pollution-related increases among individuals living in high pollution areas. Although this possibility seems remote, the counterbalancing influences of one or more covariables (such as smoking or occupational exposures) could lead to the observation of only small net effects in cancer frequency. In general, the lack of an association is not parsimoniously explained in terms of opposing influences that prevent the detection of an association.

Another important covariable is socio-economic status. The cancer rates of many sites are influenced by social class status. To isolate the effects of air pollution, this confounding variable must be taken into account. The simple method of stratifying incidence cases into fairly homogeneous groups based on median family income of the tract of residence was adopted to remove the effects of social class. This strategy allows comparison of rates "free" from differences in socio-economic status. Income is an often used

measure of social class, although in this case income is measured indirectly by using the median family income of the tract of residence rather than the personal income of each case. This "ecologic" approach to measuring social class confirmed known relationships between income and cancer incidence for many of the 38 sites. For example, Table 3 illustrates, as expected, increasing rates cancer of the lung, bronchus and trachea associated with decreasing income levels for each of the four age groups. Both positive and negative associations reported by other investigators were replicated for a variety of sites in the TNCS data with the use of median family income (not shown in this presentation). Since previously reported relationships between social class and cancer occurrence emerge with the use of tract level family income data, assurance is provided that this indirect measure indicates an individual's socio-economic status.

Nevertheless, the use of median family income to control for differences in social class is still open to debate. If cancer patients represent a random sample of the residents of a census tract, then the median family income would reflect the general level of income for those patients. Cancer cases are not a random sample, but rather tend to be older and, for many cancer sites, with lower income than the average tract levels. In other words, if median family income represents socio-economic status, then it overstates the income level of cancer patients. An alternative argument can be made that the median family income for a census tract indicates the average social and economic environment in which a person lives and should not be taken as representative of personal income. Furthermore, income correlates with many other indirect measures of the socio-economic environment of the census tract, such as median educational level, percentage of black residents, and percentage of residents older than 65 years. Thus, median family income probably serves to represent a number of dimensions of social class,[3] and employing other "ecologic" variables adds

little new information. Although most geographic studies of air pollution[3-6,9,12,13] have used indirect measures of social class, the question of whether income (measured indirectly) sufficiently reflects social status of individuals to allow for reasonable control of this fundamental confounding variable remains unanswered for this study as well as others.

The methods that have been used to interpolate air quality values over geographic areas vary from simplistic to extremely sophisticated. The method employed here using weighted geometric means of TSP and CO levels is somewhere between these extremes. The three-year average which was used undoubtedly reflects general pollution levels, and little sensitivity is gained by incorporating (modeling) other influences such as stationary sources of pollution, weather patterns, or seasonal trends. The critical issue in determining pollution exposure of an area is not how air quality is measured but when it is measured.

It is widely agreed that most cancers take 10 to 20 years to develop to clinically recognizable stages. Therefore, the measurements of air quality should more or less correspond to the onset of the cancer under study. This is not the case in this study or any other study of cancer and pollution. In general, cancer rates and air pollution levels are assessed at roughly the same time. The assumption implicit in this approach is that present air quality measurements reflect rather consistent relative differences among geographic areas over time. It is plausible but not demonstrable that the relative patterns of TSP and CO for the 737 census tracts have not changed dramatically, although the average levels of these pollutants have indeed varied over the last 10-20 years. To the degree that this assumption holds, the census tracts in this study are correctly classified with regard to air pollution exposure.

Another related assumption underlying geographic comparisons is that the populations being compared are fairly stable.

The analysis of recently measured air pollution and cancer rates assumes that the populations classified at high and low exposure are sufficiently stable so that the individuals who make up the cancer rates are those individuals who were, by and large, exposed to the differing levels of pollution. Again, this assumption is necessary for useful geographic comparisons of cancer occurrence but rarely, if ever, verified. A reasonable speculation about the Bay Area's population is that the urban areas have been somewhat stable for the last decade or two and the rural areas (a small portion of the SMSA) have experienced large amounts of residential growth.

Where the attempt to demonstrate a relationship between cancer incidence rates and urban air pollution as measured by TSP and CO shows negative results, two competing possibilities exist for not demonstrating an association: no relationship exists, or the data and methodology lack the power to demonstrate an existing relationship. That is, a verdict of "not guilty" leaves two alternatives: "innocent" or "insufficient evidence." These two possibilities are not specifically separated in this analysis. The fact that the high and low exposure groups did not have extremely different levels of pollution is the most plausible candidate for the lack of sensitivity in these data. The dose-response relationship between air quality and the occurrence of cancer is not known, making it impossible to rigorously define adequate differences in air pollution level that would lead to probable identification of differences in cancer rates. As in most epidemiologic studies, a calculation of statistical power is not realistic and leaves only informal and non-structured assessments of the "innocent" versus "insufficient evidence" issue.

REFERENCES

1. J. H. Angel, C. M. Fletcher, and I. D. Hill, "Respiratory Illness in factory and office workers," Br. J. Dis. Chest 59(1965): 66-80.

2. P. J. Lawther, R. E. Waller, and M. Henderson, "Air pollution and exacerbation of bronchitis," Thorax 25(1970): 525-539.

3. W. Winkelstein, S. Kantor, "Stomach cancer: Positive association with suspended particulate air pollution," Arch. Environ. Health 18 (1969): 544-547.

4. W. Winkelstein, S. Kantor, E. W. David, et al., "The relationship of air pollution and economic status to total mortality and selected respiratory system mortality in men," Arch. Envir. Hlth. 14(1967): 162-171.

5. W. Weiss, "Lung cancer mortality and urban air pollution," Am. J. Pub. Hlth., 68(1978): 773-775.

6. P. Stocks, "Statistics of cancer of the lung," J. Fac. Radio. 6(1955): 166-173.

7. I.T.T. Higgins. "Epidemiologic evidence on the carcinogenic risk of air pollution," INSERM 52(1976): 41-52 (IARC Scientific Publ. No. 13).

8. S. J. Cutler and J. L. Young, "Third National Cancer Survey: Incidence Data," Washington, D.C.: U.S. Department of Health, Education and Welfare (1975).

9. P. Stocks, "On the relation between atmospheric pollution in urban and rural localities and mortality from cancer, bronchitis, and pneumonia, with particular reference to 3:4 benzopyrene, beryllium, molybdenum, vanadium and arsenic," Brit. J. Cancer 14 (1960): 397-418.

10. W. W. Holland, A. E. Bennett, I. R. Cameron, et al., "Health effects of particulate pollution: Reappraising the evidence," Am. J. Epid. 110 (1979): 527-659.

11. C. M. Shy, "Epidemiologic evidence and the United States Air Quality Standards," Am. J. Epid. 110(1979): 661-671.

12. L. B. Lave and E. P. Seskin, Air Pollution and Human Health, The Johns Hopkins University Press (1977).

13. R. Mendelsohn and G. Orcutt, "An empirical analysis of air pollution dose-response curves," J. Env. Econ. and Management 6(1979): 85-106.

ACKNOWLEDGMENTS

The work described in this report was funded by the California Air Resources Board (ARB) under Contract No. A7-185-30, and by the Office of Health and Environmental Research, Assistant Secretary for Environment of the U.S. Department of Energy under Contract No. W-7405-ENG-48.

The work is part of the ongoing PARAP/PAREP (Populations at Risk to Air/Environmental Pollution) project, which was previously supported by the U.S. Environmental Protection Agency (Interagency Agreement EPA-IAG-DG-0075 between EPA and DOE).

Past and present PARAP/PAREP project managers are: Craig Hollowell and Donald M. Austin (LBL); Bill Nelson (EPA); Walter Weyzen and John Viren (DOE); and Dane Westerdahl (ARB).

Extensive use was made of the resources in SEEDIS (Socio-Economic Environmental Demographic Information System), maintained by the LBL Computer Science and Applied Mathematics (CSAM) Department.

Data used directly or indirectly in this analysis were provided by John Young of the National Cancer Institute; the National Air Data Branch of EPA; Carol Evans, Jerome Mersch and Al Wehe of EPA; Joe Ash and Carmen Benkovitz of Brookhaven National Laboratory; Phyllis Fuja of Argonne National Laboratory; and Andy Loebl and Richard Olson of Oak Ridge National Laboratory.

Table 1. List of cancer sites and Third National Cancer Survey (TNCS) codes for cancers investigated.

Cancer Sites	TNCS Codes
1. lip	400-409
2. tongue	410-419
3. salivary gland	420-429
4. gum and mouth	430-459
5. nasopharynx	471-479
6. esophagus	500-509
7. stomach	510-519
8. small intestine	520-529
9. colon	531-539
10. rectum	540-541
11. liver	550
12. gall bladder	560
13. pancreas	570-579
14. larynx	610-619
15. lung, bronchus and trachea	621,623,624
16. breast	740-749
17. cervix (invasive)	800-809
18. corpus	820
19. uterus (nos)	829
20. ovary	830
21. vulva	842
22. prostate	859
23. testes	869
24. penis	870
25. bladder	889
26. kidney	890
27. bones and joints	700-709
28. soft tissue	710-719
29. melanomas	730-739
30. eye and orbit	900-909
31. brain	910-919 except 9531-9533
32. thyroid	930

33.	Hodgkin's disease	9653-9683
34.	multiple myeloma	691
35.	acute lymphocytic leukemia	690-699, type 9825
36.	chronic lymphocytic leukemia	690-699, type 9827
37.	acute granulocytic leukemia	690-699, type 9865
38.	chronic granulocytic leukemia	690-699, type 9867

Table 2. Estimated total suspended particulate (TSP) and carbon monoxide (CO) levels (micrograms per cubic meter) for the 737 census tracts of the San Francisco-Oakland SMSA.

	geometric mean	geometric standard deviation	minimum	median	maximum	number of tracts
entire SMSA:						
TSP	47.3	5.06	37.4	47.0	77.6	417
CO	2152.2	307.73	1697.0	2008.0	2935.7	342
pollution exposure above median:						
TSP	50.8	5.75	47.0	48.3	77.6	298
CO	2376.6	274.8	2008.0	2381.8	2935.7	373
pollution exposure below median:						
TSP	44.7	2.23	37.4	45.7	47.0	715
CO	1911.8	64.0	1697.0	1926.8	2008.0	715

*Ten tracts did not meet the requirements for a "valid" estimate of air pollution, and twelve tracts were not included in the TNCS data.

Table 3. Average annual age-specific incidence rates per 100,000 for cancer of of the lung, bronchus and trachea in white males by TSP exposure (high/low) and income level (3 groups). The corresponding population at risk is given in parentheses.

Income	TSP Level (micrograms per cubic meter) below 47 mean = 44.7	above 47 mean = 50.8	all tracts mean = 47.3
age 35-44:			
<$ 9,700	24.6(8,131)	34.8(14,354)	31.1(22,485)
$9,700-$12,000	30.1(21,019)	21.0(22,216)	25.4(43,235)
>$12,000	15.7(57,146)	24.4(31,415)	24.4(88,561)
all tracts	20.1(86,296)	25.5(67,985)	22.5(154,281)
age 45-54:			
<$ 9,700	102.1(8,488)	141.6(13,889)	126.6(22,377)
$9,700-$12,000	85.3(23,831)	84.1(22,985)	84.7(46,816)
>$12,000	61.8(60.960)	64.0(29,689)	62.5(90,649)
all tracts	71.5(93,279)	87.1(66,563)	80.0(159,842)
age 55-64:			
<$ 9,700	230.3(7,670)	316.8(13,256)	285.1(20,926)
$9,700-$12,000	242.9(20,999)	262.7(18,274)	252.1(39,273)
>$12,000	181.6(39,091)	214.4(20,367)	192.9(59,458)
all tracts	206.1(67,760)	257.6(51,897)	228.4(119,657)
age over 64:			
<$ 9,700	393.7(10,413)	539.5(15,508)	480.9(25,921)
$9,700-$12,000	450.4(19,391)	409.5(17,906)	430.8(37,297)
>$12,000	365.3(27,464)	381.1(16,267)	371.2(43,731)
all tracts	399.3(57,268)	440.8(49,681)	418.6(106,949)

Table 4. The number of positive associations among the twelve comparisons (tied values excluded) between TSP levels and average annual age-specific cancer incidence rates for 38 sites with significance probabilities ("p-values") for males and females. The number of comparisons for a given sex and site is less than 12 when deaths did not occur in all age-income categories.

	MALE			FEMALE		
CANCER SITE	No.of cases	No.of positive associations	p-value	No.of cases	No.of positive associations	p-value
1. lip	106	6/12	.613	13	3/5	.500
2. tongue	137	3/11	.967	70	7/11	.274
3. salivary gland	66	5/11	.726	92	5/11	.726
4. gum and mouth	168	6/12	.613	111	6/12	.613
5. nasopharynx	33	5/11	.726	16	6/9	.254
6. esophagus	181	7/11	.274	109	7/10	.172
7. stomach	567	5/12	.806	376	9/12	.073
8. small intestine	60	7/10	.172	45	4/8	.637
9. colon	1280	7/12	.387	1532	6/12	.613
10. rectum	730	8/12	.194	661	6/12	.613
11. liver	123	7/11	.274	64	4/11	.887
12. gall bladder	31	4/9	.746	76	8/11	.113
13. pancreas	467	11/12	.003	372	6/10	.377
14. larynx	312	8/12	.194	51	6/11	.500
15. lung,bronch,trachea	2641	9/12	.073	1630	4/12	.927
16. breast	30	8/11	.113	3755	4/12	.927
17. cervix (invasive)	-	-	-	475	7/12	.387
18. corpus	-	-	-	1268	4/12	.927
19. uterus (nos)	-	-	-	74	8/12	.194
20. ovary	-	-	-	651	10/12	.019
21. vulva	-	-	-	69	6/11	.500
22. prostate	1985	5/11	.726	-	-	-
23. testes	89	9/12	.073	-	-	-

24.	penis	32	7/12	.387	-	-	-
25.	bladder	862	5/12	.806	339	8/12	.194
26.	kidney	277	6/12	.613	146	6/12	.613
27.	bones & joints	18	4/8	.637	20	5/7	.227
28.	soft tissue	81	8/12	.194	57	6/12	.613
29.	melanoma	164	6/12	.613	186	4/12	.927
30.	eye & orbit	23	5/11	.726	29	6/11	.500
31.	brain	212	4/12	.927	198	6/12	.613
32.	thyroid	98	7/12	.387	206	6/12	.613
33.	Hodgkin's disease	104	5/12	.806	63	6/12	.613
34.	multiple myeloma	160	5/11	.726	139	6/12	.613
35.	acute lymph.leukemia	12	2/5	.813	10	3/5	.500
36.	chron.lymph.leukemia	262	5/9	.500	65	2/7	.938
37.	acute gran. leukemia	97	4/12	.927	77	6/12	.613
38.	chron.gran. leukemia	44	4/11	.887	27	3/9	.910
	total, all sites	11452	197/367	.087	13072	199/375	.128

Table 5. The number of positive associations among the twelve comparisons (tied values excluded) between CO levels and average annual age-specific cancer incidence rates for 38 sites with significance probabilities ("p-values") for males and females. The number of comparisons for a given sex and site is less than 12 when deaths did not occur in all age-income categories.

	MALE			FEMALE		
CANCER SITE	No.of cases	No.of positive associations	p-value	No.of cases	No.of positive associations	p-value
1. lip	106	6/12	.613	13	4/5	.188
2. tongue	137	8/11	.113	70	5/11	.726
3. salivary gland	66	5/11	.726	92	4/11	.887
4. gum and mouth	168	4/12	.927	111	4/12	.927
5. nasopharynx	33	6/11	.500	16	3/9	.910
6. esophagus	181	5/11	.726	109	6/10	.377
7. stomach	567	6/12	.613	376	8/12	.194
8. small intestine	60	4/10	.828	45	5/8	.363
9. colon	1280	6/12	.613	1532	6/12	.613
10. rectum	730	6/12	.613	661	7/12	.387
11. liver	123	6/11	.500	64	8/11	.113
12. gall bladder	31	2/9	.980	76	5/11	.726
13. pancreas	467	5/12	.806	372	3/10	.945
14. larynx	312	3/12	.981	51	5/11	.726
15. lung,bronch,trachea	2641	7/12	.387	1630	4/12	.927
16. breast	30	2/11	.994	3755	4/12	.927
17. cervix (invasive)	-	-	-	475	4/12	.927
18. corpus	-	-	-	1268	5/12	.806
19. uterus (nos)	-	-	-	74	7/12	.387
20. ovary	-	-	-	651	5/12	.806
21. vulva	-	-	-	69	4/11	.887
22. prostate	1985	6/11	.500	-	-	-
23. testes	89	3/12	.981	-	-	-

24. penis	32	5/12	.806	-	-	-
25. bladder	862	5/12	.806	339	4/12	.927
26. kidney	277	5/12	.806	146	7/12	.387
27. bones & joints	18	3/8	.855	20	6/7	.063
28. soft tissue	81	7/12	.387	57	6/12	.613
29. melanoma	164	5/12	.806	186	4/12	.927
30. eye & orbit	23	4/11	.887	29	6/11	.500
31. brain	212	5/12	.806	198	6/12	.613
32. thyroid	98	4/12	.927	206	6/12	.613
33. Hodgkin's disease	104	8/12	.194	63	9/12	.073
34. multiple myeloma	160	7/11	.274	139	8/12	.194
35. acute lymph.leukemia	12	3/5	.500	10	2/5	.813
36. chron.lymph.leukemia	262	8/9	.020	65	2/7	.938
37. acute gran. leukemia	97	8/12	.194	77	5/12	.806
38. chron.gran. leukemia	44	8/11	.113	27	4/9	.746
total, all sites	11452	175/367	.826	13072	181/375	.765

FIGURES

Figure 12. Distribution of six income-exposure (CO) categories for the 737 census tracts of the San Francisco-Oakland SMSA.

Figure 13. Association between total suspended particulate and cancer incidence in standard normal deviations, for males, females and both sexes combined.

Figure 14. Association between carbon monoxide and cancer incidence in standard normal deviations, for males, females and both sexes combined.

This report was done with support from the Department of Energy. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the Department of Energy.

Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

TECHNICAL INFORMATION DEPARTMENT
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720