PB84-199785

Epidemiological Study of the Incidence of Cancer as Related to Industrial Emissions in Contra Costa County, California

California Dept. of Health Services, Emeryville
Resource for Cancer Epidemiology Section

Prepared for

Health Effects Research Lab.
Research Triangle Park, NC

Jun 84

U.S. DEPARTMENT OF COMMERCE
National Technical Information Service
NTIS®

PB84-199785

EPA-600/1-84-008
June 1984

EPIDEMIOLOGICAL STUDY OF THE INCIDENCE OF CANCER AS RELATED TO INDUSTRIAL EMISSIONS IN CONTRA COSTA COUNTY, CALIFORNIA

by

Donald F. Austin
Verne E. Nelson
Bix E. Swain
Linda F. Johnson
California Department of Health Services
Resource for Cancer Epidemiology Section
Emeryville, CA 94608

Grant No. R806396-01

Project Officer

Wilson R. Riggan
Population Studies Division
Health Effects Research Laboratory
Research Triangle Park, NC 27711

HEALTH EFFECTS RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
CINCINNATI OH 45268

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-600/1-84-008		PB84 199785

4. TITLE AND SUBTITLE	5. REPORT DATE
Epidemiological Study of the Incidence of Cancer as Related to Industrial Emissions in Contra Costa County, California	June 1984
	6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
D.F. Austin, M.D., V. Nelson, B. Swain, L. Johnson, S. Lum, P. Flessel	

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT NO.
Dept. of Health Services, Resource for Cancer Epidemiology Section and California Tumor Registry, 5850 Shellmound St., Suite 200, Emeryville, CA 94608	L104
	11. CONTRACT/GRANT NO. R806396

12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Health Effects Research Laboratory U.S. Environmental Protection Agency Research Triangle Park, N.C. 27711	Final: Sept. 1978 to Dec. 1982
	14. SPONSORING AGENCY CODE EPA/600/11

15. SUPPLEMENTARY NOTES

Project Officer: Wilson B. Riggan, Ph.D., BD, HERL

16. ABSTRACT

The purpose of this study was to examine the relationship of lung cancer incidence in Contra Costa County to ambient levels of air pollution. It was suspected that the presence of heavy industry in the county, mainly petrochemical plants and oil refineries, could be a contributing factor. Initially, an incidence analysis established that the industrial portion of the county had an excess of lung cancer as compared to the remaining non-industrial portion. Air pollution patterns were subsequently determined by five permanent air monitoring stations and ten temporary stations which monitored the levels of 12 air pollutants for a period of one year. By correlating the 1970-79 lung cancer rates for each census tract and tract levels of air pollution constituents, a statistically significant relationship between ambient air SO_4 and lung cancer in males, but not in females, was found. However, when adjusted for the percent of the working population categorized as blue collar, the association was eliminated. An interview study of 249 cases and 373 controls was then conducted. Demographic, work history, residential history, dietary, and smoking history questions comprised the bulk of the data collected. Analysis indicated that the major contribution to lung cancer in the county was due to cigarette smoking. No significant association between lung cancer risk and measured constituents of air pollution was found. Of five broad occupational categories (indicating possible hazardous exposures) none had any significant relationship to lung cancer.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS (This Report)	21. NO. OF PAGES
RELEASE TO PUBLIC	UNCLASSIFIED	79
	20. SECURITY CLASS (This page) UNCLASSIFIED	22. PRICE

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

FOREWORD

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

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

This epidemiologic study assesses the risk of industrial emissions on the health of the residents of Contra Costa County. Concern about the health effects of exposure to air pollution generated by the heavily industrialized area of the county prompted this investigation. Initially the study focussed on cancer of the trachea, lung or bronchus, comparing the Industrial and Non-industrial areas of the county and reviewed incidence cases and rates spanning the years 1975-1979. Subsequently, air monitoring of the industrial emissions at 15 sampling stations provided data used herein to calculate census tract specific air pollutant measures for a correlation analysis of air pollutants and census tract specific lung cancer incidence. Finally, an assessment of the risk of lung cancer for county residents was conducted through a case-control questionnaire study linked to census tract specific air pollution measurements.

F. Gordon Hueter, Ph.D.
Director
Health Effects Research Laboratory

ABSTRACT

This study of the relationship of lung cancer incidence in Contra Costa County to ambient levels of air pollution was generated by the concern of public officials and private citizen groups concerned about reports of elevated lung cancer incidence in the county. It had been suspected by some that the presence of industrial plants in the county, mainly petrochemical refineries, could be a contributing factor. The study consisted of five parts.

First, an incidence analysis established that when the county was divided into two parts, the Industrial area of the county had an excess of lung cancer as compared to the remaining Non-industrial area.

More detailed information on the patterns of air pollution in the county were obtained in the second phase of the study. Five permanent air monitoring stations and ten temporary stations monitored the levels of 12 air pollutants for a period of one year. These data were incorporated into later phases of the study.

In the third portion of the study, through a correlation analysis of 1970-79 lung cancer rates and various air pollution constituents, a relationship between ambient air SO_4 and lung cancer in white males, but not in white females, was found to be statistically significant. However, the percent of the working population categorized as blue-collar was also associated with lung cancer in white males and the previous association between lung cancer in white males and ambient air SO_4 levels was eliminated when this third factor was taken into consideration.

Part four of the study was to have consisted of a linkage of occupational group cohorts to registry cancer incidence files but was not conducted for lack of easy availability of occupational group records.

Part five of the study was an analysis of case-control interview data on a final sample of 622 individuals. Demographic, work history, residential history, dietary, and smoking history information comprised the bulk of the data collected.

Analysis of the data indicated that the major contribution to lung cancer in Contra Costa County was due to cigarette smoking. Further, there was no identified effect on lung cancer risk contributed by any measured constituent of air pollution. Of five broad occupational categories (indicating possible hazardous exposures) none had any significant relationship to lung cancer.

CONTENTS

FIGURES

TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

ABBREVIATIONS

AIHL	-- Air and Industrial Hygiene Laboratory
BAA	-- benz(a)anthracene
BAAQMD	-- Bay Area Air Quality Management District
BAP	-- benzo(a)pyrene
BGP	-- benzo(ghi)perylene
BSO	-- benzene soluble organics
CHR	-- chrysene
COR	-- coronene
CTR	-- California Tumor Registry
EPA	-- United States Environmental Protection Agency
ICD-O	-- International Classification of Diseases for Oncology
IIO	-- Index of Industries and Occupations
$\mu g/m^3$	-- micrograms per cubic meter
NCI	-- National Cancer Institute
ng/m^3	-- nanograms per cubic meter
NIOSH	-- National Institute of Occupational Safety and Health
OR	-- odds ratio
p	-- significance probability
PAH	-- polycyclic aromatic hydrocarbons
RCE	-- Resource for Cancer Epidemiology
RDD	-- random digit dialing
rev/m^3	-- revertants per cubic meter
S9	-- metabolic activator for mutagenic tester strains
SEER	-- Surveillance, Epidemiology and End Results Program
SF-O SMSA	-- San Francisco-Oakland Standard Metropolitan Statistical Area
SIR	-- standardized incidence ratios
T98	-- a mutagenic tester strain
TSP	-- total suspended particulate

SYMBOLS

NO_3	-- nitrate
Pb	-- lead
SO_2	-- sulphur dioxide gas
SO_4	-- sulphate

ACKNOWLEDGMENTS

This study could not have been carried out without the advice, guidance and continued support of a great number of people. To properly recognize each of these people is not possible in such a report. However, certain individuals must be singled out for special acknowledgements.

The guidance and support of Dr. Wilson R. Riggan, our United States Environmental Protection Agency (EPA) Project Officer, who provided helpful advice regarding the administration, funding and scientific review of the project is gratefully acknowledged. The professional consultation of several reknowned specialists provided invaluable technical guidance for the project staff. These persons include: Dr. William Gaffey, Statistical Consultant, formerly of Stanford Research Institute International; Mr. James Sandberg, Meteorologist, Bay Area Air Quality Management District (BAAQMD); Dr. Warren Winkelstein, Jr., former Dean, School of Public Health, University of California, Berkeley; Dr. Richard Brand, Professor, School of Public Health, University of California, Berkeley; Dr. Alice Whittemore, Adjunct Professor, Department of Family, Community and Preventive Medicine, Stanford University, and Dr. Byron Wm. Brown, Professor and Head, Division of Biostatistics, Stanford University.

The collaboration of field and laboratory staff of excellent air pollution analytic laboratories under the direction of Mr. Dario Levaggi, Executive Director, BAAQMD, and Dr. Jerome Wesolowski, Chief, Air and Industrial Hygiene Laboratory, (AIHL) Berkeley, and the employment of the Ames test for mutagenicity by Dr. Peter Flessel, Air and Industrial Hygiene Laboratory, contributed significantly to the scientific excellence of the physical and biological measurements used in this study.

The numerous consultations provided by technical staff of the National Institute of Occupational Safety and Health (NIOSH) are gratefully acknowledged.

The efforts of a number of members of the Resource for Cancer Epidemiology Section (RCE) who collected and coded the cancer incidence data under a National Cancer Institute (NCI) contract are acknowledged, although their individual mention would be prohibitively long; likewise, the efforts of a large staff of highly trained interviewers and supervisors, directed by Dr. Vonnie Gurgin and Ms. Mary Hauck, who carried out the nearly 700 lengthy and difficult interviews used in this study. The efforts of former RCE staff, including Dr. William Mandel and Mildred Snyder are recognized. The active collaboration and contributions of Dr. Eva Glazer and Maggie Chiang are gratefully acknowledged.

The analysis of the enormous amount of data gathered under the auspices of this project were carried out by a rather large analytic staff, each contributing in different areas. Key contributions by staff are acknowledged in the following areas: Susan Lum - computer mapping of air pollution data; Lynn Palmer - incidence rates; Margaret Hackett - review and consolidation of cancer case incidence file; Susan Brown - pre-analysis edits, match/merging of census data in correlation analysis; and Pamela Chamberlin - census tract coding, tabular presentations, initial incidence report organization.

A separate note of appreciation is due to Kim Ikeno and members of her clerical unit for the management of myriad office tasks related to the conduct of this project, including the preparation of this final report.

Lastly, it is necessary to recognize the efforts of a large number of people not formally associated with this project. Prominent among them are former Senator John A. Nejedly and Assemblyman William Campbell, both of Contra Costa County, whose personal and legislative efforts kept this project operating. Others who played significant roles include members of citizens' and environmental groups, the press, concerned members of state and county government and many citizens of Contra Costa County.

SECTION 1

OVERVIEW

INTRODUCTION

Contra Costa County, located in the northeastern part of the San Francisco Bay Area, is one of an aggregate of 39 U.S. counties found to have a high mortality rate for specific cancer sites (Blot et al., 1977). The fact that the county also has a section which is heavily industrialized, with five major petroleum refineries and numerous petrochemical plants, and the fact that 68% of the total stationary air pollution in the Bay Area originates from the county (AIHL, 1980a) prompted an epidemiological study of the incidence of cancer in Contra Costa County funded by the United States Environmental Protection Agency. The major objective was to determine whether industrial emissions have a measurable effect on cancer occurrence. The study was originally proposed by and funds granted to the Contra Costa County Health Department. Following its award, the grant was rejected by the Contra Costa County Board of Supervisors and returned to the EPA with the request that the study be carried out by the California Department of Health Services. This request was acceptable to both the EPA and the State of California and the grant was subsequently awarded to the California Department of Health Services and carried out by the Resource for Cancer Epidemiology Section of that agency. The study was to have consisted of five parts:

1. A comparison of the cancer incidence in the heavily industrialized sections of Contra Costa County to that in the remainder of the county.

2. Air monitoring, consisting of sampling and chemical analysis of air to determine the levels of particulate pollution components in the ambient air.

3. Correlation analysis of lung cancer incidence rates with air pollution constituents and census tract characteristics.

4. Occupational group monitoring to detect occupational groups at elevated risk of cancer.

5. A case-control study to identify specific environmental factors responsible for any excessive amount of cancer incidence in Contra Costa County.

DATA SOURCE BACKGROUND

Since 1972, the Resource for Cancer Epidemiology has maintained a cancer surveillance system (Linden and Austin, 1974) in the San Francisco-Oakland Standard Metropolitan Statistical Area (SF-O-SMSA), a five county area that includes Contra Costa County and which has a base population of over three million. The RCE is under contract to the National Cancer Institute to provide cancer incidence data for the NCI Surveillance, Epidemiology and End Results Program (SEER) (NCI, 1981). An earlier program of the NCI, the Third National Cancer Survey (1969-1971), also conducted by the RCE in these same five counties allowed those data also to be used (NCI, 1975) in the analyses which follow.

The RCE cancer data files are maintained by a sophisticated, automated consolidated cancer information system developed in cooperation with NCI, and serving as a model for other U.S. cancer registries.

SECTION 2

INCIDENCE ANALYSIS

INTRODUCTION

A preliminary analysis of cancer incidence in Contra Costa County for the period 1972-1975 for all sites combined, lung and bronchus, trachea, stomach, prostate and lymphoma was completed prior to this analysis for the Non-industrial area and Industrial area of the county. A line drawn by subjective means separated the approximate Industrial area of the county from the rest of the county, constituting the Non-industrial area. The results of the preliminary analysis showed that a significant difference existed only for cancer of the lung and bronchus. Therefore, these analyses were limited to cancer of the lung and bronchus.

METHODS

Lung Cancer Case Selection

Lung cancer cases included for incidence analysis were malignant, invasive, resident incidence cases for the primary sites of lung, bronchus and trachea for the period of 1969-1978. Extreme effort was used to eliminate any erroneous case allocation such as duplicate reports, incorrect dates of diagnosis, improper census tract assignments, and other factors such as incorrect demographic and temporal information.

Industrial, Non-industrial Area Definition

In addition, a more objective allocation of Industrial and Non-industrial census tracts was done. Tracts which were both zoned and used for heavy industry were assigned to the Industrial area with the remaining census tracts in the county comprising the Non-industrial area (Figure 1). Comparison of the lung and bronchus cancer incidence in the two areas for both sexes and white and black races were conducted, as were temporal trends for race and sex categories in each area and for the county as a whole.

Development of Population Estimates

Population estimates for 1970-78 by age, race and sex were generated from census tract data using the 1970 census, a 1975 special census of the county, and the 1980 census counts of persons by race by using standard demographic techniques of interpolation and extrapolation. Aspects of the

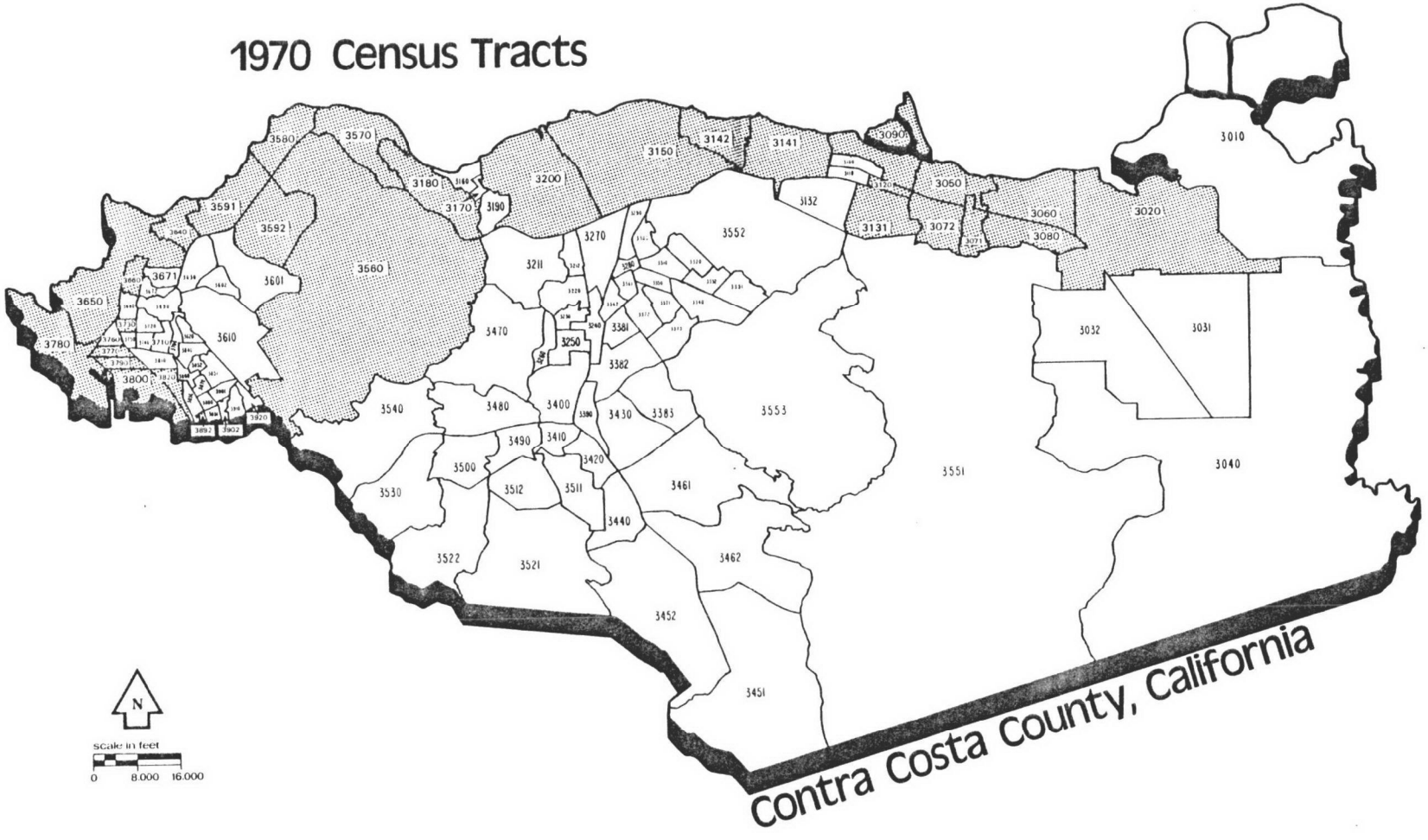

Figure 1. 1970 census tracts, Contra Costa County.

population denominator file which could possibly contribute to erroneous results were also ruled out.

DATA ANALYSIS

Lung cancer incidence rates for the Industrial and Non-industrial areas of Contra Costa County show a widening difference for each race and sex over the period of 1969-1978. Final incidence rates for the period 1975-1979 show a significant excess of lung cancer in the Industrial area residents (Table 1). Using the total rate figures of Table 1, the lung cancer rate for the Industrial area (69.4) exceeds that of the Non-industrial area (50.0) by 38.8% or approximately 40%.

Both the number of new lung cancer cases and the number of deaths increased faster than the population in Contra Costa County during the period 1969-1979 (Table 2). The number of cases more than doubled during the period 1969 to 1977. During the last two years the number of cases appears to have decreased slightly.

Lung cancer incidence reporting by Industrial/Non-industrial area of residence is presented in Table 3 for individual years 1969-1979. While roughly progressive increases in lung cancer incidence occur for the Non-industrial area, a peak in cases occurs in the Industrial area in 1975 followed by a gradual reduction in later years.

Though Table 4 demonstrates a numerical difference in the number of lung cancer cases among black males and females in the Industrial/Non-industrial areas, a statistically significant difference in rates could not be demonstrated (see Table 1). The black population is concentrated in two areas of the county and primarily resides in the Industrial area. Therefore, the black population would have nearly identical exposures to air pollution and would make analyses of lung cancer incidence based on differences in pollutant exposures difficult. For the above reasons, the study focused on lung cancer incidence in the white population.

Figure 2 presents lung cancer rate data for white males for the Industrial and Non-industrial areas for the period of 1970-1978. A difference in incidence rates occurs for the period with a maximum in 1976. The observed differences were statistically significant at the .01 level in 1972-1976 period and the .05 level in 1977 and 1978 as Table 5 illustrates.

In Figure 3, a trend of high white female rates is indicated in the Industrial area but for fewer years than in the male comparison. Elevated rates are present in the 1974-1978 period. However, statistical significance (at the .01 level) occurs only in the shorter 1975-1977 period. As in the white male comparison, a decline in rates occurs in the last two years (Table 5).

In Table 6, white males in the Industrial area had higher age-specific rates in every age group except the two youngest 0-29 years and 30-34 years. These differences were statistically significant (at the .05 level) for ages 50-54 and 55-59, using the difference of means test.

TABLE 1. FIVE-YEAR AVERAGE ANNUAL AGE-ADJUSTED* INCIDENCE RATES FOR CANCER OF THE BRONCHUS AND LUNG, INDUSTRIAL AND NON-INDUSTRIAL AREAS OF CONTRA COSTA COUNTY, 1975-79

Group	Industrial area	Non-industrial area
Males	101.2 (89.4 - 113.0)**	76.9 (71.2 - 82.6)**
Females	41.7 (34.6 - 48.8)	30.3 (27.2 - 33.4)
White males	108.6 (94.5 - 122.7)	77.9 (71.8 - 84.0)
White females	44.7 (36.5 - 52.9)	31.0 (27.7 - 34.3)
Black males	92.9 (65.9 - 120.0)	70.7 (40.7 -100.7)
Black females	37.2 (21.5 - 52.9)	15.6 (3.8 - 27.4)
Total	69.4 (62.7 - 76.1)	50.0 (47.1 - 52.9)

*Rates are expressed as cases per 100,000 population and are adjusted to the 1970 U.S. standard.

**Numbers in parentheses are 95% confidence intervals.

TABLE 2. NUMBER OF INCIDENCE CASES OF LUNG CANCER CASES*, CANCER DEATHS**, AND ESTIMATED POPULATION***, CONTRA COSTA COUNTY, WHITE AND BLACK POPULATION, 1969-1979

Year	Number of cases	Number of deaths	Estimated population
1969	176	131	555,083
1970	222	165	559,491
1971	236	182	563,909
1972	239	178	568,328
1973	255	177	572,735
1974	271	201	577,148
1975	282	214	581,640
1976	303	203	598,830
1977	359	250	614,171
1978	322	260	629,511
1979	333	233	644,852

*Resident incidence cases with site codes 162.0-162.9 by ICD-0 (World Health Organization, 1976).

**Deaths among residents with primary cause coded 162 by ICD-0 (World Health Organization, 1976).

***Population estimated using 1970 and 1980 U.S. census totals and Contra Costa County census of 1975.

Source: Unpublished SEER data, Resource for Cancer Epidemiology Section

TABLE 3. NUMBER OF LUNG CANCER CASES* BY SEX AND YEAR, INDUSTRIAL AND NON-INDUSTRIAL AREAS, CONTRA COSTA COUNTY, WHITE POPULATION, 1969-1979

	Industrial area			Non-industrial area		
Year	Total	White male	White female	Total	White male	White female
1969-1979	663	481	182	1985	1387	598
1969	41	33	8	109	83	26
1970	45	32	13	152	115	37
1971	46	32	14	166	123	43
1972	62	52	10	158	119	39
1973	56	47	9	171	114	57
1974	53	41	12	190	144	46
1975	81	55	26	174	121	53
1976	74	49	25	187	126	61
1977	76	52	24	239	153	86
1978	65	39	26	210	142	68
1979	64	49	15	229	147	82

*Excludes 99 cases for the period 1969-79 where a census tract number could not be assigned to the address.

Source: Unpublished SEER data, Resource for Cancer Epidemiology Section

TABLE 4. NUMBER OF LUNG CANCER CASES* BY SEX AND YEAR, INDUSTRIAL AND NON-INDUSTRIAL AREAS, CONTRA COSTA COUNTY, BLACK POPULATION, 1969-1979

Year	Industrial			Non-Industrial		
	Total	Black male	Black female	Total	Black male	Black female
1969-1979	128	94	34	58	44	14
1969	7	6	1	5	5	0
1970	7	5	2	4	4	0
1971	11	7	4	3	1	2
1972	10	8	2	2	1	1
1973	10	9	1	9	6	3
1974	12	10	2	4	3	1
1975	12	9	3	7	5	2
1976	15	12	3	6	4	2
1977	12	8	4	4	4	0
1978	16	10	6	9	7	2
1979	16	10	6	5	4	1

*Excludes 6 cases for the period 1969-79 where a census tract number could not be assigned to the address.

Source: Unpublished SEER data, Resource for Cancer Epidemiology Section

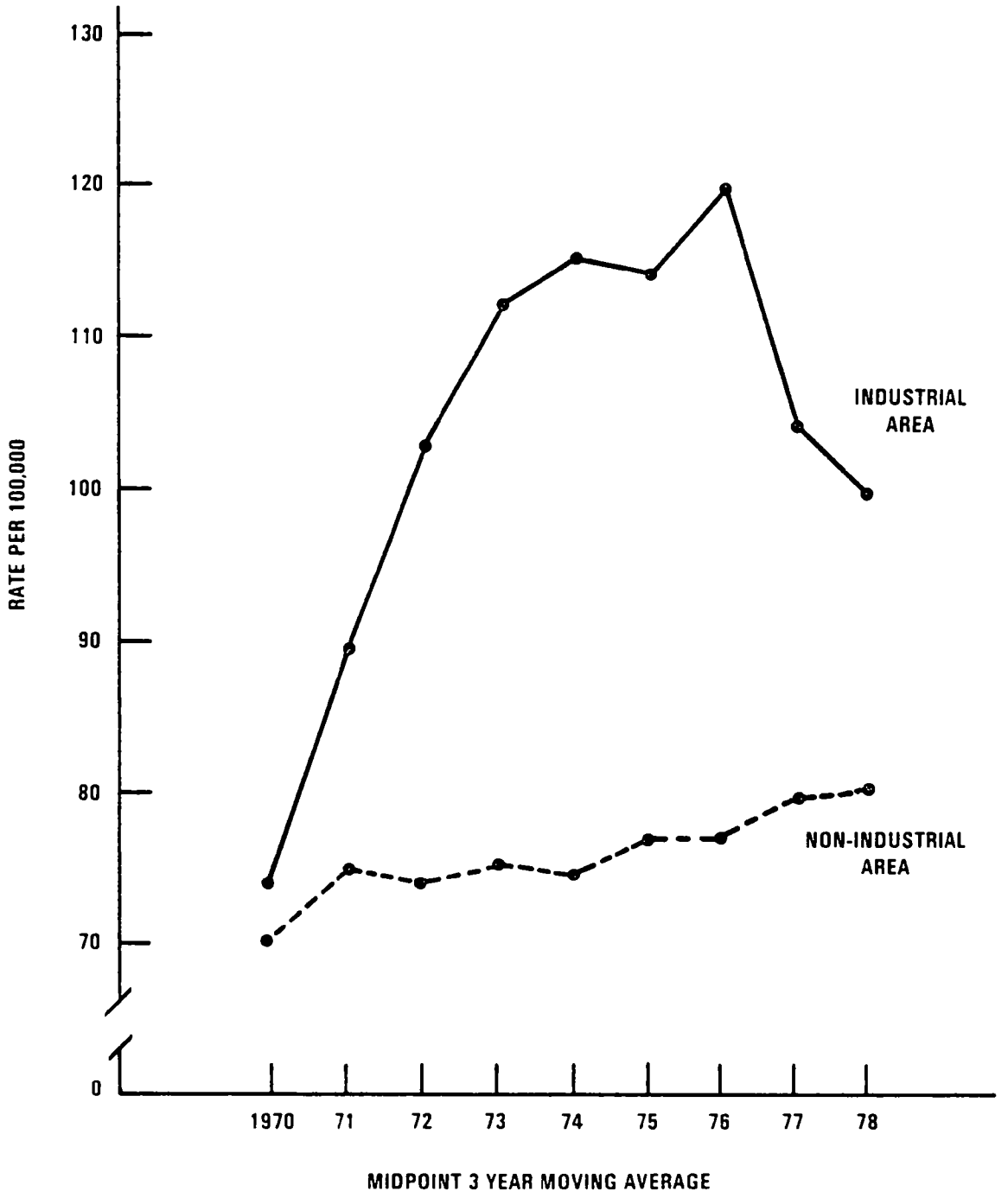

Figure 2. Three year annual age adjusted lung cancer incidence rates. Industrial and Non-industrial areas. Contra Costa County, white males, 1970-1978.

TABLE 5. THREE YEAR AVERAGE AGE ADJUSTED RATES FOR LUNG CANCER INDUSTRIAL AND NON-INDUSTRIAL AREAS*, CONTRA COSTA COUNTY, WHITE MALES AND WHITE FEMALES, 1970-1978

Three year average	Average age-adjusted incidence rates per 100,000 population			
	White males		White females	
	Industrial area	Non-industrial area	Industrial area	Non-industrial area
1970	74.2	70.4	24.9	18.1
1971	89.7	75.9	26.6	20.2
1972	103.0	74.3	23.9	23.3
1973	112.3	76.5	21.4	23.1
1974	115.1	75.7	32.5	24.8
1975	114.6	78.3	42.7	24.8
1976	120.3	78.3	50.3	30.4
1977	104.2	79.4	48.1	31.8
1978	100.0	79.9	40.0	33.9

*Excludes 99 cases where a census tract could not be assigned to the patient's address.

Source: Unpublished SEER data, Resource for Cancer Epidemiology Section

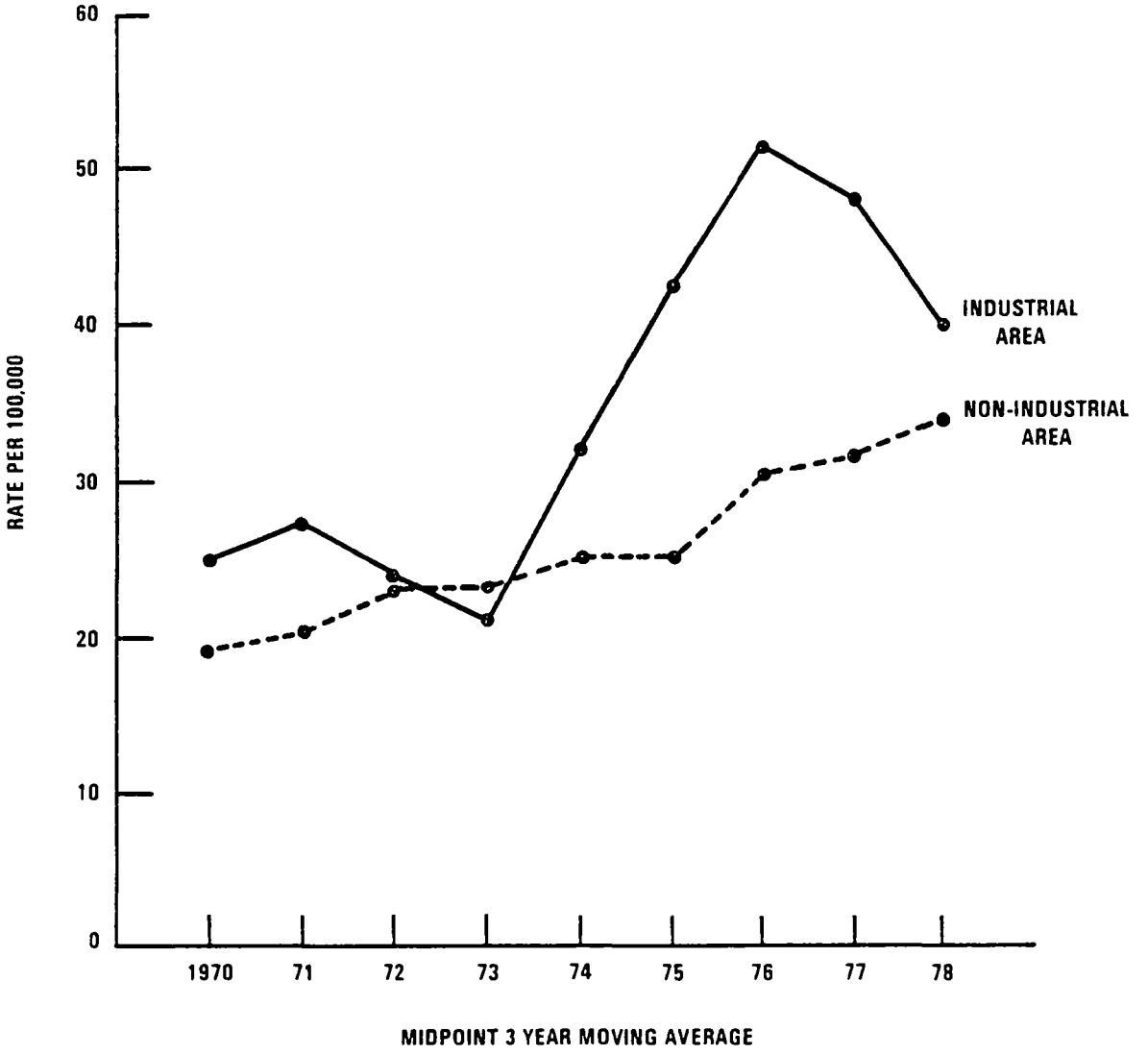

Figure 3. Three year annual age adjusted lung cancer incidence rates. Industrial and Non-industrial areas. Contra Costa County, white females, 1970-1978.

TABLE 6. AGE SPECIFIC INCIDENCE RATES* FOR LUNG CANCER, INDUSTRIAL AND NON-INDUSTRIAL AREAS, CONTRA COSTA COUNTY WHITE MALES AND FEMALES, 1975-1979

Age group	White males		White females	
	Industrial	Non-industrial	Industrial	Non-industrial
0-29	-	0.2	0.7	0.4
30-34	-	3.6	7.3	2.3
35-44	19.9	11.9	15.6	8.8
45-49	80.6	52.6	90.6	28.7
50-54	201.4	61.8	70.5	55.8
55-59	324.8	147.1	173.0	90.7
60-64	374.0	303.6	192.2	166.5
65-69	421.3	395.0	167.8	139.9
70-74	605.0	525.7	195.1	143.4
75+	724.4	586.4	137.5	125.6

*Rates expressed as cases per 100,000 population

Source: Unpublished SEER data, Resource for Cancer Epidemiology Section

White females in the Industrial area had higher rates in every group when compared to the Non-industrial area. The difference in rates for females in the 45-49 and 55-59 age groups was statistically significant at the .05 level (Table 6).

Figure 4 presents cumulative risk comparison for the 1975-1979 period. For both white males and white females the major Industrial/Non-industrial difference in the risk is established by age 60. By age 75, the cumulative risk of acquiring lung cancer for white males in the Industrial area is 10.6% versus 7.6% for males in the Non-industrial area, equating to an excess risk in the Industrial area of approximately 40%. Comparable risk figures for white females by age 75 are 4.6% in the Industrial area versus 3.2% in the Non-industrial area equating to an excess risk in the Industrial area of approximately 44% as shown in Table 7.

CONCLUSIONS

A review of lung cancer incidence data by age, race and sex for various time periods between 1969 and 1978 shows excess risk in the Industrial area of Contra Costa County. Overall the excess is approximately 40% in the time period, with a maximum difference occurring around 1976, followed by narrowing differences in 1977 and 1978.

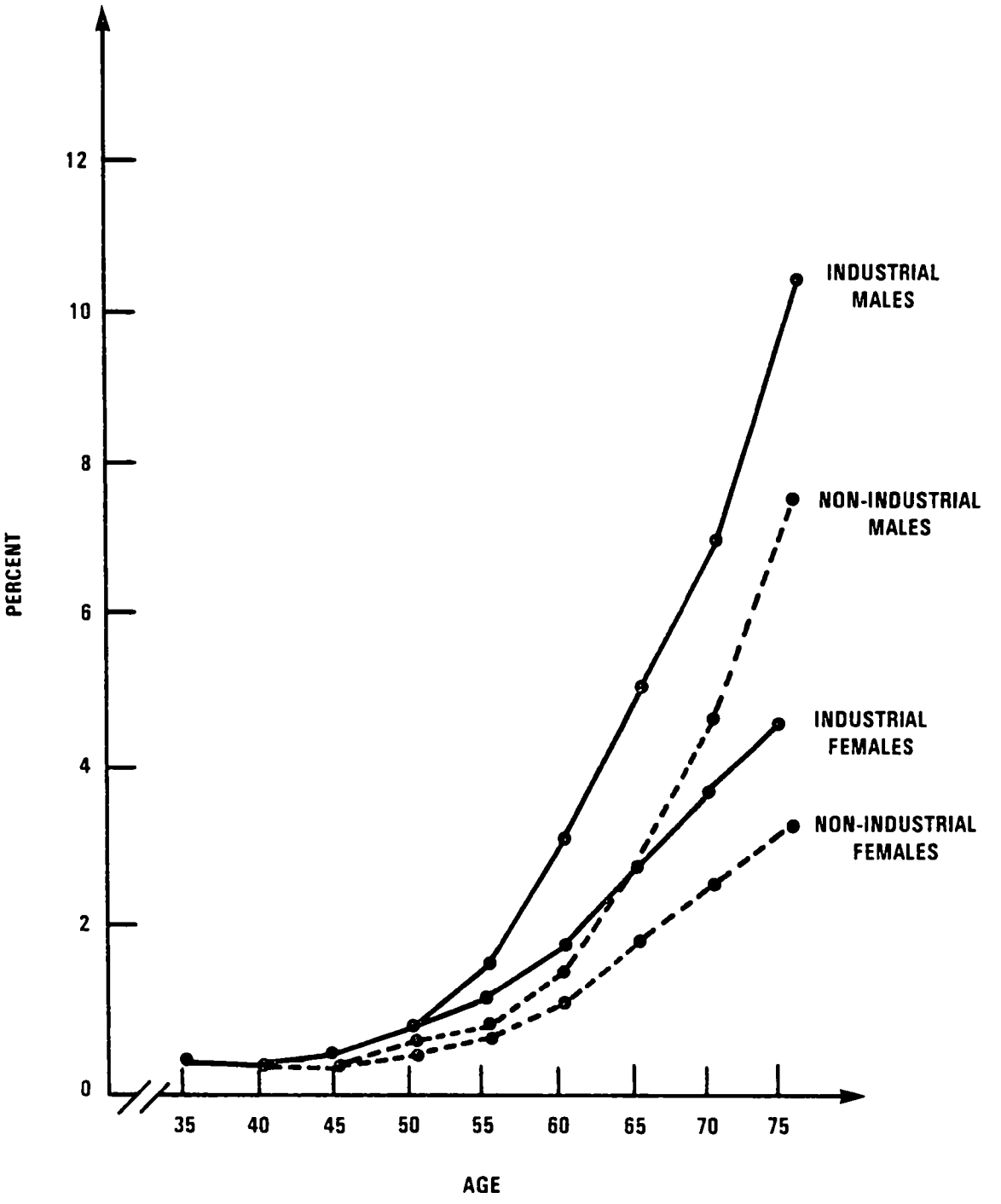

Figure 4. Contra Costa County incidence analysis. Cumulative risk of acquiring lung cancer. White population, 1975-1979.

TABLE 7. CUMULATIVE RISK OF ACQUIRING LUNG CANCER EXPRESSED AS PERCENT, AGE 30-75, INDUSTRIAL AND NON-INDUSTRIAL AREAS, CONTRA COSTA COUNTY, WHITE MALES AND WHITE FEMALES, 1975-1979

Age	White males		White females	
	Industrial area	Non-industrial area	Industrial area	Non-industrial area
35	-	-	.1	-
40	-	.1	.1	-
45	.2	.1	.2	.1
50	.6	.4	.6	.2
55	1.6	.7	1.0	.5
60	3.2	1.4	1.8	1.0
65	5.1	3.0	2.8	1.8
70	7.2	4.9	3.6	2.5
75	10.6	7.6	4.6	3.2

Source: Unpublished SEER data, Resource for Cancer Epidemiology Section

SECTION 3

AIR POLLUTION MONITORING

INTRODUCTION

The air pollution study had a number of goals:

1. To provide data for determining the association, or lack thereof, of the present cancer incidence rates with air pollution.

2. To determine whether or not mutagenic activity as measured by the Ames assay could be accounted for by the chemical characterization of the samples.

3. To develop baseline data on ambient air pollutants for comparisons with future measurements and for use in future epidemiological cancer studies.

METHODS

Equipment, Placement and Specimen Collection

A total of 15 hi-volume particulate samplers were strategically sited at 13 locations in Contra Costa County and two locations in adjacent counties so as to characterize air quality variations over the entire county (Figure 5).

Air particulate material was collected on 8" x 10" glass-fiber filters (EPA Grade Whatman) in standard high-volume samplers which were collected every sixth day at each of the 15 sampling stations from November, 1978, to October, 1979. After sample collection, the filters were weighed to determine the amounts of total suspended particulate material and delivered to the Air and Industrial Hygiene Laboratory on a weekly basis. There the filters were logged in, cut and the pieces distributed for further analysis.

Particulate matter was analyzed for total suspended particulate (TSP), benzene soluble organics (BSO), sulfate (SO_4), nitrate (NO_3), lead (Pb), selected polycyclic aromatic hydrocarbons (PAH), and mutagenic activity. Because of the constraints of sample size and resources, the measurements of PAH and mutagenicity could not be done on each sample. Instead, samples were composited over the three natural meterological seasons of the San Francisco Bay Area, viz. Winter: November to February, Spring: March to

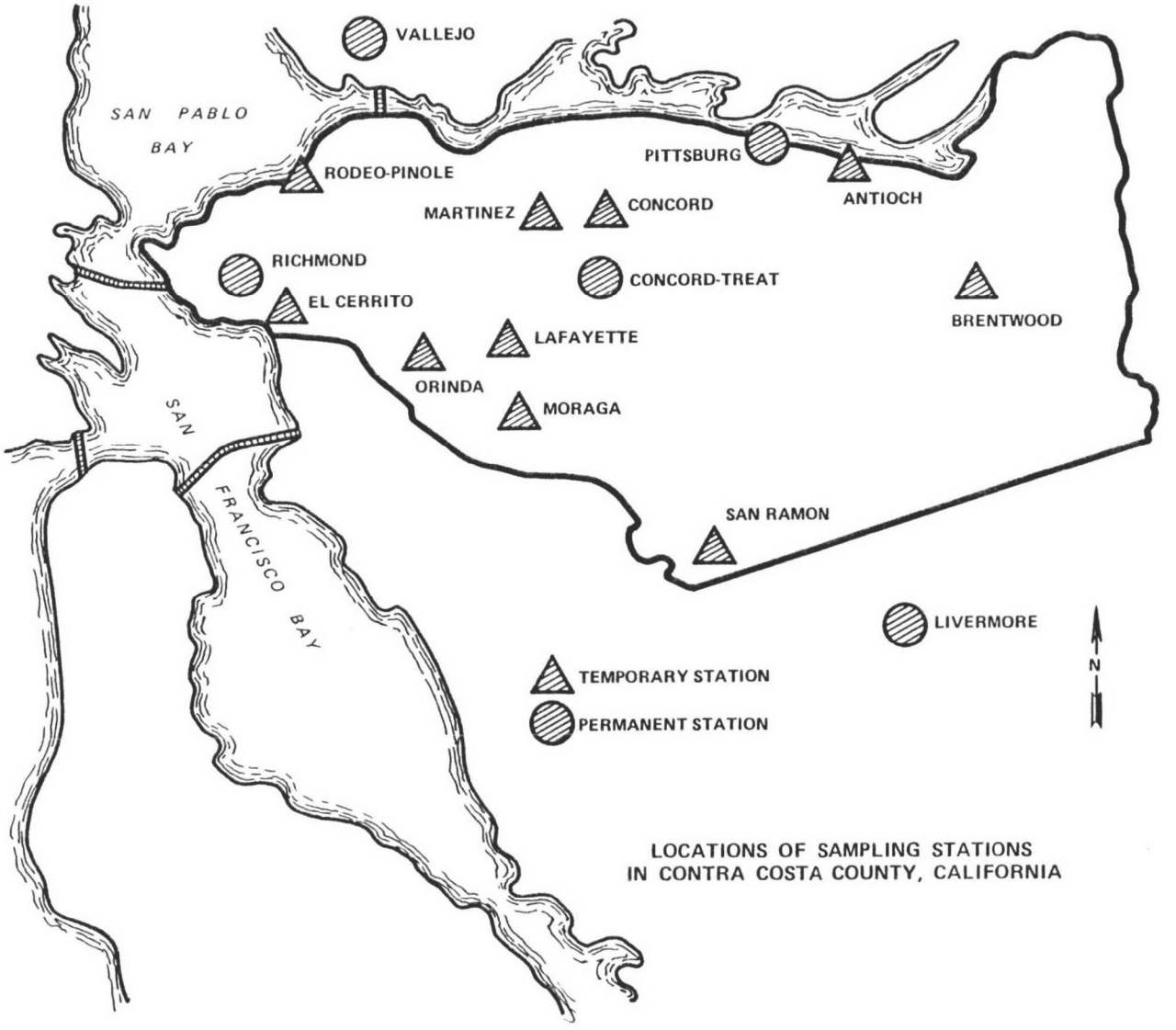

Figure 5. Locations of air monitoring sampling stations in Contra Costa County

June, and Summer: July to October. For each of the 15 stations, composites were then prepared by combining all the samples collected over the three four-month periods.

Chemical and Biological Analysis

TSP, NO_3, SO_4 BSO, Pb

Standard methods were used to analyze for the following five pollutants: TSP was determined gravimetrically (BAAQMD, 1977); NO_3 colorimetrically (AIHL, 1980a); SO_4 turbidmetrically (AIHL, 1980b); BSO by Soxhlet extraction (AIHL, 1975); and Pb by wavelength dispersive X-ray fluorescence (Moore, 1976).

PAH

Specific PAH were separated by high-performance liquid chromatography (Flessel et al., 1981). Column effluents were quantitated using ultraviolet absorption and fluorescence. Five PAH were measured: benzo(a)pyrene(BAP), benzo(ghi)perylene(BGP), benz(a)anthracene(BAA), chrysene(CHR) and coronene(COR), a possible tracer for vehicle emissions.

Ames Test for Mutagenic Activity

Mutagenicity was measured using the Ames test (Ames, 1975). The Ames assay was applied according to a two-part protocol. Part one involved screening the sample in the five standard Ames tester strains both with and without metabolic activation (S9). The data gave a quantitative estimate of the mutagenic activity, the most sensitive strain, and the optimum conditions of metabolic activation for a subsequent quantative analysis.

For the initial screening, five tester strains were used. All samples showed activity in at least one stain and generally the most mutagenic activity was measured in one specfic strain (T98). In part two of the assay, all composite samples were then analyzed using that strain and dose response curves were obtained. A measure of mutagenicity was obtained from the slope of the curves. A typical dose-response curve is shown in Figure 6.

DATA ANALYSIS

Overview

The results of analysis of air particulate material for five standard pollutants (TSP, BSO, Pb, SO_4, and NO_3) are summarized in Table 8. The median and maximum value for each pollutant are listed for winter, spring and summer. The median and maximum levels of these pollutants were generally the highest in the winter, reflecting the occurrence of meterological inversions. The highest level for benzene-soluble organics (BSO) measured was 41.3 $\mu g/m^3$ in Concord in December, 1978, and the lowest was less than 0.8 $\mu g/m^3$ at several sites in January and February, 1979. The highest Pb concentration was found in Antioch, also in December, while the highest TSP concentration was in Brentwood in October.

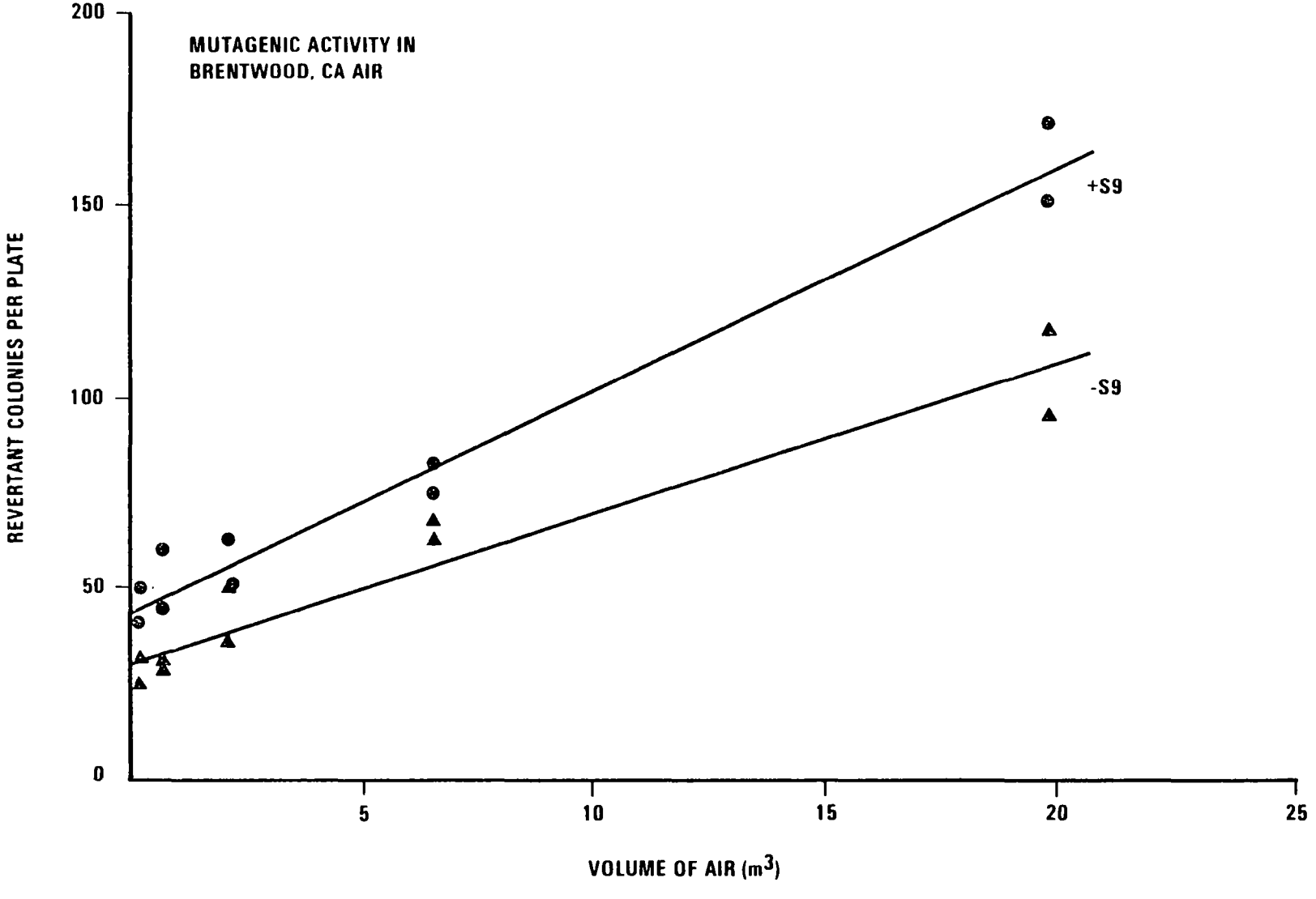

Figure 6. Dose-response curves for a composite sample collected at Brentwood, Ca., November, 1978-February, 1979.

TABLE 8. ANALYSIS OF AIR PARTICULATE MATERIAL BY SEASON, CONTRA COSTA COUNTY, NOVEMBER 1978 - OCTOBER 1979

Pollutant	Twenty-four hour median value* $\mu g/m^3$			Twenty-four hour maximum value $\mu g/m^3$		
	Nov 78 - Feb 79	Mar - Jun 79	July - Oct 79	Nov 78 - Feb 79	Mar - Jun 79	July - Oct 79
Total suspended particulate material	60	42	52	229	126	418
Benzene soluble organics	4.8	1.4	2.1	41.3	32.1	8.1
Lead	0.7	0.2	0.27	2.5	0.6	1.2
SO_4	6.9	5.3	6.3	19.3	15.9	13.2
NO_3	7.2	3.1	4.0	31.6	16.6	15.2

*Median values were calculated from results obtained on samples collected every six days at the fifteen sampling stations.

Relation Between PAH and Mutagenicity

The correlation coefficient (Spearman) between mutagenicity values for each of the monitoring stations and the concentrations of the five PAH measured (BAP, BAA, CHR, BGP, and COR) were 0.33, 0.41, 0.36, 0.31, and 0.51, respectively, indicating that chemical quantitation of the PAH was not a good measure of the mutagenicity of the air. This is consistent with the observation that BAP, BAA, BGP, CHR, and COR represent only about 2% of the total mutagenic activity of the air. This was demonstrated by comparing the mutagenicity of the five PAH in amounts proportional to their average levels measured in the ambient winter air composites with the average mutagenicity measured in the actual ambient air samples suggesting a poor association between the ambient air chemical and biological measurements. This indicated that a more chemical characterization of ambient air is needed in order to account for the mutagencity as measured by the Ames test.

Geographic Distribution of Air Pollution

In order to correlate cancer incidence data to air pollution measurements, information about the variations in levels of air pollution by census tract was needed. Contour maps showing the geographic distribution of the levels of the seven measured pollutants were constructed using a computer program called SYMAP (Harvard University Laboratory for Computer Graphics and Spatial Analysis, 1975). Sampling station coordinates and associated pollutant levels were used to construct a matrix containing the pollution levels throughout the county. Contours were then constructed from the matrix values (Figures 7,8).

Values of the pollutants from each of the 15 sampling stations obtained were used to compute estimated values for each of the population centroids for each of the 115 census tracts in the county. The Pearson product moment correlation coefficients between pollutants were computed for the 113 cenus tracts (two atypical tracts, a naval base and a retirement community, were removed). The correlation coefficients between pollutants for the 113 census tracts show very similar relationships to those based on the 15 monitoring stations. While the correlation between Pb and BSO for the 15 monitoring stations was 0.78, the value for the 113 census tracts was 0.70 (Table 9). Likewise, the correlation between BSO and BAA (one of the five PAH) for the 15 monitoring stations was 0.87, the value for the 113 census tracts was 0.89. For the fifteen stations, the correlation between BSO and the two mutagenicity tests (with and without S9) were 0.28 and 0.24 while the correlations based on the 113 census tracts were 0.39 and 0.28, respectively. Thus, it was felt that the real relationships between various air pollution constituents, as evidenced from the correlation coefficients with the data from the 15 stations, were preserved by the computer mapping technique so that the computed values of the 113 census tracts could be used with confidence in subsequent analyses using cancer data.

Figures 7-8 show contour maps for Pb, SO_4, BSO, the five PAH and mutagenicity. The geographic distributions for Pb and BSO are similar in showing that the highest levels are found in a north-south band located in central Costa Costa County, a region corresponding to the Diablo Valley, a

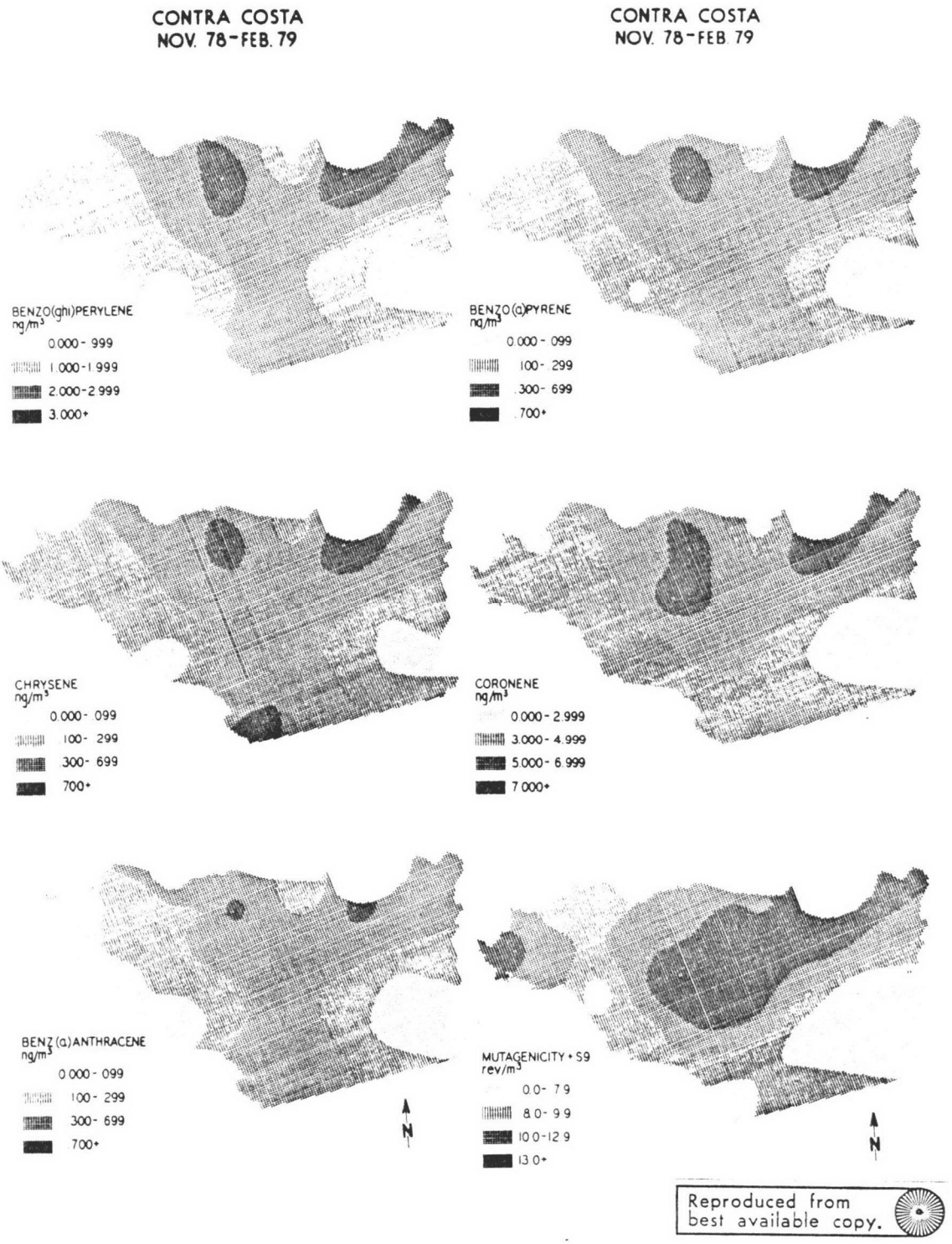

Figure 7. Contour map of indicated pollutants, Contra Costa County, November 1978-February, 1979

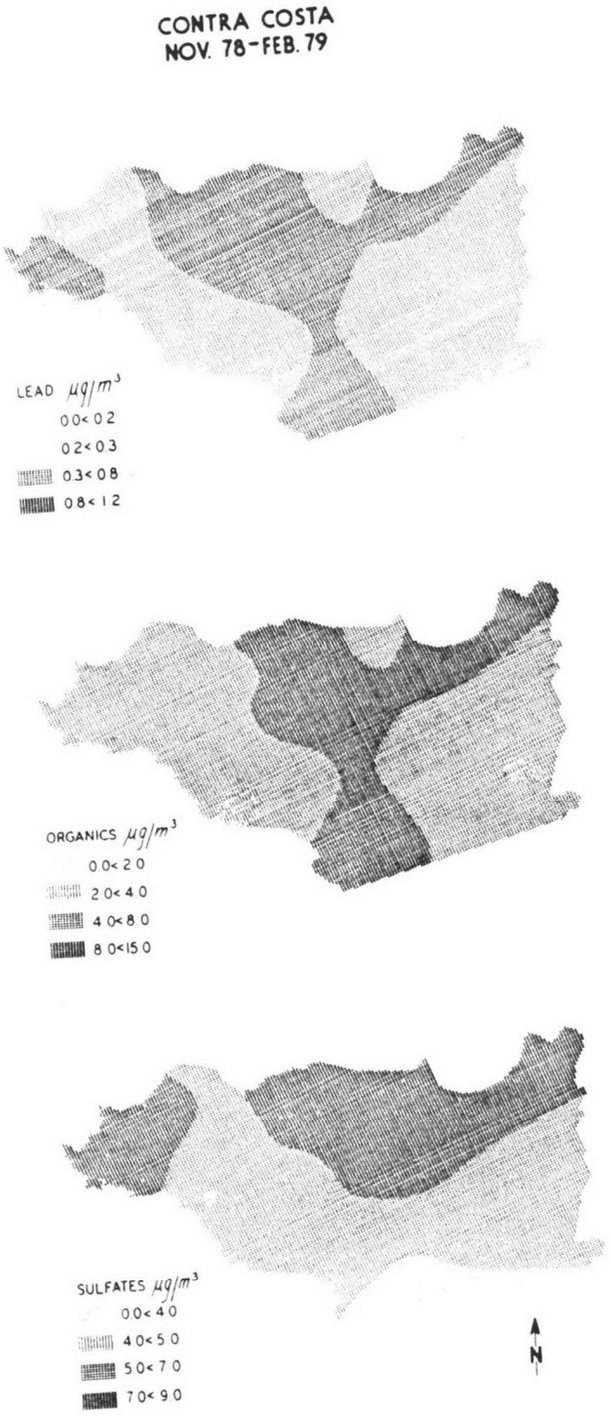

Figure 8. Contour map of indicated pollutants, Contra Costa County, November 1978-February, 1979

TABLE 9. PEARSON CORRELATION COEFFICIENTS FOR THE MEAN ANNUAL MEASURED VALUES FOR 15 MONITORING STATIONS AND FOR COMPUTED VALUES FOR 113 CENSUS TRACT CENTROIDS, CONTRA COSTA COUNTY, NOVEMBER 1978 - OCTOBER 1979

Comparison	15 Stations	113 Census tracts
BSO vs Pb	.78	.70
BSO vs Mut (-S9)	.24	.28
BSO vs Mut (+S9)	.28	.39
BSO vs BAA	.87	.89
BSO vs BAP	.80	.84
BSO vs BGP	.64	.71
BSO vs CHR	.83	.84
BSO vs SO_4	.19	.17
TSP vs NO_3	.64	.75

natural pollution sink and the path of a freeway. The Pb map is consistent with the fact that the largest source of Pb in the area is the automobile. Comparison of the BSO and Pb maps suggests the contribution of the automobile to the BSO levels may be significant. The SO_4 distribution differs from the Pb distribution by running in an east-west direction along the industrial belt. This is consistent with the fact that sulphur dioxide gas (SO_2), the precursor of SO_4, is emitted by stationary sources, primarily chemical industries, refineries and power plants, all located along the industrial belt. The patterns of the five PAH are similar to one another and to lead.

Sources of PAH and Mutagenicity

Table 10 gives the correlation coefficients among the pollutants measured. Values are high between Pb, BSO, and the five PAH, implicating vehicles as the source of these pollutants. However, the correlation coefficients between mutagenicity and Pb, BSO, and the five PAH are small suggesting that mutagenicity concentrations may come from multiple sources.

Information can be obtained on sources by the use of simple ratio techniques. The ratio, for example, between BAP and BGP for a number of combustion sources have been established. Automobiles tend to have the lowest ratio, 0.3 to 0.4 while other sources tend to be equal or greater than 1. The ratio found in winter in Contra Costa County ranged from 0.08 to 0.25, consistent with the automobile being a major contributor of BAP. The high correlation between Pb and PAH, the similarity of Pb and PAH contour maps, and the value of the ratio all point to the automobile as a major contributor to PAH.

CONCLUSIONS

The Ames test can be carried out with enough precision to make it a useful tool for determining trends, geographic distributions, or temporal variations of the mutagenic potential of ambient air particulate matter.

It is advantageous to use both the Ames test and chemical characterization together in attempting to predict the potential carcinogenic effects of ambient air particulate matter.

In order to account for the mutagenicity as measured by the Ames test, a more complete chemical characterization of ambient air particulate matter will be needed.

In Contra Costa County, mobile sources are significant contributors to PAH; however, more research is needed to define the major sources of particulate mutagens.

TABLE 10. SPEARMAN RANK CORRELATION COEFFICIENTS BETWEEN MEASURED POLLUTANTS FOR CONTRA COSTA COUNTY, NOVEMBER, 1978 - FEBRUARY, 1979

	Pb	BSO	BAP	BAA	CHR	BGP	COR	+S9	-S9	TSP	SO_4	NO_3
Pb	1.00											
BSO	.87	1.00										
BAP	.84	.92	1.00									
BAA	.84	.94	.90	1.00								
CHR	.82	.96	.92	.94	1.00							
BGP	.88	.87	.93	.92	.88	1.00						
COR	.87	.82	.88	.82	.80	.91	1.00					
+S9	.44	.51	.33	.41	.36	.31	.51	1.00				
-S9	.46	.59	.39	.44	.47	.34	.42	.88	1.00			
TSP	.33	.58	.39	.65	.59	.48	.35	.45	.48	1.00		
SO_4	.39	.26	.31	.20	.31	.37	.47	.27	.25	.07	1.00	
NO_3	.14	.20	-.04	.23	.21	.12	.06	.39	.42	.71	.38	1.00

SECTION 4

CORRELATION ANALYSIS

INTRODUCTION

A correlation analysis between cancer incidence rates with air pollution constituents and census tract characteristics was conducted to establish any relationships which might suggest a cause of the excess of lung cancer found in the Industrial area residents and to formulate hypotheses which could be tested through the case-control study.

METHODS

Case Selection, Population Estimates and Socio-economic Variables

The following information was available for the correlation analysis:

1. Estimates of air pollution data.

2. Refined population estimates by age, race and sex.

3. Number of lung/bronchus cancer cases by age, race and sex.

4. Demographic information (i.e. socio-economic variables) obtained from the 1970 U.S. census and the 1975 Contra Costa County census.

5. Ten and five-year average annual age-adjusted lung cancer incidence rates, by sex and race based on (2) and (3) above.

Correlation Methods

Pearson product moment correlation coefficients were computed for comparisons between estimated annual values for each air pollutant constituent for each census tract and the 5- and 10-year average annual age-adjusted incidence rates for cancer of the lung for males and females for each census tract. Certain census tract-specific socio-economic variables from the 1970 United States census and 1975 Contra Costa County census were also correlated with cancer incidence and air pollutants. Partial correlations were computed to control for these attributes.

TABLE 11. CORRELATION OF THE CENSUS TRACT SPECIFIC, FIVE-YEAR AVERAGE ANNUAL AGE-ADJUSTED LUNG CANCER INCIDENCE RATES IN CONTRA COSTA COUNTY FOR *WHITE MALES AND WHITE FEMALES**, 1970-74 AND 1975-79

	White males 1970-74	White males 1975-79	White females 1970-74	White females 1975-79
White males 1970-74	1.00	-	-	-
White males 1975-79	.17	1.00	-	-
White females 1970-74	- .02	.14	1.00	-
White females 1975-79	.23	.54 ($p<0.01$)	.05	1.00

*Correlation of the 10-year rates for white males vs white females = .50 ($p<0.0001$) (N = 113 census tracts)

DATA ANALYSIS

Since the correlation of census tract-specific average annual lung cancer rates for the 5-year periods, 1970-1974 and 1975-1979, for white males and white females differed significantly from zero for only two out of the six possible correlations, only ten-year rates were used (Table 11). No statistically significant differences were demonstrated for the black population owing possibly to the sparse geographical distribution of the black population mentioned in the incidence analysis. The correlation between the ten-year census tract-specific lung cancer incidence rates for white males and females was 0.33 ($p<.0001$) for all tracts and 0.50 ($p<.0001$) with the two atypical tracts removed (a naval base and a retirement community).

The ten-year census tract-specific lung cancer incidence rates were significantly correlated with only one measure of particulate air pollution, that of SO_4 (Table 12). The correlation coefficient for white males was 0.46 ($p<.0001$) and for white females the correlation coefficient was 0.16, not significant. Controlling for the percent of households in the census tract in which the head of the household had resided in the unit for 20 plus years, as determined in the 1975 Contra Costa Census, did not effect the correlation. Controlling for the percent of census tract residents of Spanish origin reduced the correlation slightly for males (Table 13). Controlling for the percent of blue collar workers significantly reduced the correlation between SO_4 and lung cancer for males. This reduction was

TABLE 12. CORRELATION OF THE TEN-YEAR (1970-79) AVERAGE ANNUAL AGE-ADJUSTED LUNG CANCER INCIDENCE RATE TO PARTICULATE AIR POLLUTION CONSTITUENTS* BY INDIVIDUAL CENSUS TRACT**, CONTRA COSTA COUNTY, CALIFORNIA

Air pollution constituent	vs	Average annual incidence rate (per 10^5)	
		White males	White females
Mutagenicity (-S9)		-.01	-.07
Mutagenicity (+S9)		.05	-.01
BSO		.11	-.09
BAA		-.05	-.18
BAP		-.09	-.18
BGP		-.13	-.16
CHR		-.03	-.14
TSP		.20 (P<0.05)	-.07
P_b		.11	.06
NO_3		-.09	-.24 (P<0.05)
SO_4		.46 (P<0.0001)	.16

*Mean annual value, November 1978 - October 1979

**113 tracts

TABLE 13. CORRELATION BETWEEN SULFATES AND THE 1970-79 TEN-YEAR AVERAGE ANNUAL LUNG CANCER INCIDENCE RATES, BY INDIVIDUAL CENSUS TRACT*, CONTRA COSTA COUNTY, CONTROLLING FOR SEVERAL CENSUS TRACT CHARACTERISTICS**

Controlling for:	SO_4 vs White males	SO_4 vs White females
1. Nothing	.46 (p<.0001)	.16
2. Head of household resident of census tract for 20+ Years	.46 (p<.0001)	.13
3. Percent Spanish origin	.42 (p<0.001)	.19 (p<0.05)
4. Percent blue-collar workers***	.21 (p<0.05)	.02
a) Percent unskilled laborer***	.29 (p<0.01)	.04
b) Percent skilled Laborer***	.29 (p<0.01)	.09
5. Percent of households below poverty level	.30 (p<0.001)	.07
6. Median family income	.21 (p<0.05)	-.03
7. Median school completed by head of household	.17	-.06

*113 census tracts

**Data from 1975 Contra Costa County special census

***Data from 1970 U.S. census

from independent effects of both the skilled and unskilled laborer components of blue collar workers. Controlling for the percent of households below the poverty level reduced the SO_4 lung cancer correlation only moderately. Controlling for education or income variables reduced or destroyed the correlations for males (Table 13).

CONCLUSIONS

Several conclusions can be drawn from the correlation analysis in association with the data from the incidence analysis and air monitoring portion of the study.

It is apparent that an excess of lung cancer exists in the Industrial area of the county. This excess has developed over the past decade.

Mutagenic activity, as measured by the Ames test, is identifiable in the particulate matter in ambient air in Contra Costa County and is best associated with the distribution of BSO and Pb, suggesting mostly automobile sources.

The ten-year occurrence of lung cancer in white males is only weakly but significantly associated with the distribution of SO_4 particulate matter in the ambient air. This association is reduced equally by controlling for the percent of skilled, or unskilled, laborers in each census tract. Controlling for both factors combined nearly destroys the observed association.

The source of SO_4 in the air is known to be almost exclusively from oil refining, chemical manufacture or oil combustion for electrical power generation. The association of lung cancer in males with this factor might suggest a carcinogenic effect from past petrochemical emissions were it not for three factors:

1. The current mutagenic activity in the particulates is distributed differently.

2. The association does not occur in females.

3. The association appears to be mediated through a residential pattern of blue collar workers which correlates well with the distribution of SO_4 (0.67; $p<0001$). Although these data permit no conclusions regarding the causes of lung cancer in the county, they suggest that a significant contribution to incidence by blue collar workers residing in areas of close proximity to petrochemical plants.

SECTION 5

OCCUPATIONAL MONITORING

INTRODUCTION

The objective of this project was to determine the identification of occupational cohorts at increased risks of cancer.

METHODS

A data base was available from an occupational surveillance system, based on files from the RCE first prepared in 1976, consisting of membership rosters of various unions in the San Francisco-Oakland SMSA. A cohort, composed of approximately 6,400 union members who were residents of the San Francisco-Oakland SMSA, was selected from these data. Six occupational groups were represented; asbestos workers, bakers, painters, plasterers, plumbers and roofers.

The basic methodological approach used in this study consisted of computer matching the cohort of occupational workers to the master file of cancer cases maintained by the California Tumor Registry (CTR). The observed cases for each cohort and primary site were then compared to expected cancer cases which were calculated using age-sex-year specific incidence rates for the SMSA. Standardized Incidence Ratios (SIR's) were also estimated.

DATA ANALYSIS

Increased cancer incidence was demonstrated among asbestos workers, an occupation widely recognized as having high risk for respiratory and gastrointestinal cancers.

CONCLUSIONS

This part of the study was not continued due to the lack of cooperation of additional sources of cohorts, i.e., employees and unions. Since this effort was consuming an inordinate amount of project resources it was discontinued.

SECTION 6

CASE CONTROL STUDY

INTRODUCTION

The Relationship of Lung Cancer Incidence in Contra Costa County to Air Pollution

The over-all goal of this part of the study was to identify specific environmental factors responsible for lung cancer in Contra Costa County.

The particular objectives were twofold: first, to identify any significant environmental factors which contribute to lung cancer incidence in that county and second, to specifically evaluate whether air pollutants, released into the ambient air, have had any effect on the observed incidence of lung cancer in the community.

Preceding components to this project generated findings to be addressed by this component. The incidence analysis established that when the county was divided into two parts, the Industrial area of the county had an excess of lung cancer as compared to the remaining Non-industrial area. Furthermore, this excess had increased significantly during the decade of the 1970's to a peak in 1976, after which a decline in the excess was apparent. Through a census tract specific analysis it was further demonstrated that, although an incidence rate excess of approximately 40% was found in the Industrial area of the county, considerable variability existed in both areas, so that significantly high and significantly low rates of lung cancer were found in census tracts in both areas. Thus the Industrial area was a mixture of census tracts of high, medium, and low rates, the average of which exceeded the overall rate of census tracts comprising the Non-industrial area.

Through a correlation analysis of 1970-79 lung cancer rates by census tract and various air pollution constituents only one statistically significant relationship was found to exist. That relationship was between ambient air SO_4 and lung cancer in males, but not in females. However, the percent of the population categorized as blue collar workers was also associated with lung cancer in males and the previous association between lung cancer in males and ambient air SO_4 levels was considerably reduced when this third factor was taken into consideration.

The results of the indirect studies such as the correlation analysis summarized above, led to the case-control study of individuals described below.

Literature Review

Respiratory cancer rates vary with geographic location but are still increasing over time (Doll,1981). For the period 1973-1977 the age-adjusted incidence rate in U.S. males was 78.3/100,000 for cancer of the lung and bronchus. In females, the corresponding rate was 22.2/100,000.

In the San Francisco Bay area, the age-adjusted incidence rates for lung cancer during this period were higher; 83.2/100,000 in males and 29.8/100,000 in females (Young,1981).

Overall survival rates for lung cancer are so poor that incidence and mortality rates are often used interchangeably. In the United States and most other developed countries cancer of the lung presently is responsible for almost 40 percent of male cancer mortality (del Regato,1977). Yet in 1923, the Massachusetts age-adjusted mortality rate/100,000 was 1.55 for males and 1.50 for females (Lombard,1968). The increase, which has taken place mainly during the last forty years, has been related to cigarette smoking (Hammond,1975).

U.S. men began smoking cigarettes in appreciable numbers at the turn of the century. The habit became more widespread after World War I, and reached a peak of 65 percent in 1937 (Beamis,1975;Burbank,1972). It began to decline in 1957 to about 42 percent in 1970 and 39 percent in 1975 (National Clearinghouse,1975). For women, the rise began during World War II, reaching a high of 39 percent in 1961, then falling to 30 percent by 1970 and 29 percent in 1975 (Beamis,1975). In male heavy smokers the risk for lung cancer has been established at about ten times that for non-smokers (Hammond,1975). Third National Cancer Survey interview data indicate that for women who smoke heavily, the risk may be as high as sixteen times that for non-smokers. Cigar and pipe smokers have an approximately twofold risk of developing lung cancer (Williams,1977).

Burbank (1972), using an estimated latent period of thirty years between first exposure to smoking and tumor development, has fitted both male and female lung cancer mortality rates from 1950-1968 onto a tobacco dose response curve based on estimates of past cigarette consumption.

There has also been considerable speculation on the extent to which general air pollution contributes to lung cancer mortality in urban areas. Clemmesen (1977) has suggested that it may have a non-specific effect on the bronchial mucosa to make it more susceptible to carcinogens such as tobacco smoke. Many polyaromatic hydrocarbons (PAH) are known to be carcinogenic and the most commonly used indicator for PAH is benzo(a)pyrene concentration (Pike, 1975) a known carcinogen directly measured in the present project.

More recently, Vena (1982) reported a study of the effects of air pollution in Erie County, Pennsylvania, using total suspended particulates

as a measure. He found evidence suggesting a relationship between air pollution and smoking and possibly between air pollution, occupation and smoking, for 50 or more years of exposure to air pollution.

Evidence on other pollution components is scanty. Higgins (1977) found a correlation between sulfates and lung cancer in fifty SMSA's. Blot and Fraumeni (1975), also in a correlative study, found excess lung cancer in counties that had arsenical air pollution from non-ferrous metal smelters and refineries.

METHODS

Case-Control Study Design

Case-control studies are a standard tool of epidemiological research. Simply put, a group of persons having a disease (cases in this study, lung cancer) is compared with a group of persons not having the disease (controls). In matched case-control studies, the controls are carefully selected so that they are similar to cases in important background variables. In this study, cases were matched to controls of the same race and sex, and similar age. The study populations of cases and controls are interviewed to obtain detailed information on study variables pertaining to each participant. The data are then subjected to statistical analyses to detect significant differences in known or suspected risk factors for contracting the disease. The outcome of these analyses is the product of the study. The report typically identifies risk factors which predict the probability of any individual in the study being a case. When these risk factors fulfill the criteria for causal factors their relative importance as a cause of the disease under study can be directly determined.

Questionnaire Design

The questionnaire was developed to conform to standard survey research practice. The questions were designed to be read verbatim and were extensively pretested to eliminate any ambiguities. Question order, probes, and skip patterns were organized so that the same instrument would be administered to each respondent. Care was taken to format the questionnaire for ease of administration, in the interest of both accuracy and of the time required to record the answers. The questionnaire was pilot-tested on a small sample of respiratory cancer patients and healthy residents of Alameda and Contra Costa counties who were not included in the study.

A manual of detailed interviewer instructions was prepared. This included question-by-question specifications for the questionnaire.

To ensure the accuracy of the data collection and to minimize refusals by cases and controls, we employed interviewers, trained in survey research interviewing techniques, for administration of the questionnaire and selection of controls. An interviewer training session for the study was held before the start of the field work and at intervals during the study period. This included mock interviewing during the briefing and practice interviews in the field. These were reviewed by the field work supervisor

and, where indicated, further training was conducted.

Questionnaires were edited and reviewed by the supervisor before being coded and sent for key data entry.

Case/Proxy Selection

All cases of cancer of the trachea, bronchus or lung among black or white residents of Contra Costa County, newly diagnosed between May 8, 1980 and July 31, 1981, and who were at least 35 years of age and less than 75 years of age at diagnosis, comprised a group of 332 eligible cases.

Some eligible lung cancer patients were too ill to participate in the study or, in other instances, had succumbed to their disease before an interview could be conducted. Whenever possible, a proxy for the case was identified for interviewing. Most proxies were the spouse or other immediate family member of the case but in a few instances the proxy was a close friend of the case.

Control Selection

To meet a matching requirement of at least one control per case, the required number of controls in each of 32 age, race, and sex strata was computed based on the percent distribution of the cases. The control population was therefore a representative sample of that non-cancer population of Contra Costa County with an age, race, and sex distribution equivalent to the lung cancer cases. To assure an adequate number of controls, each stratum was slightly overfilled to allow for possible subsequent deletions necessitated by insufficient respondent-supplied detail.

A target number of 350 controls for the cases was selected. Because so many variables were under consideration, each with a different prevalence in the study population, it is not possible to present a single "beta error" or a minimum required sample size. For most variables, a sample of 250 cases and 350 controls is adequate to detect relative risks of two or greater, at a significance level of = 0.05 and a probability of detection of 80% or greater (Oliphant, 1981).

Random Digit Dialing

Controls were selected from the general population of Contra Costa County by a random digit dialing (RDD) technique. Briefly, this technique required obtaining from the telephone company the telephone prefixes in use in the area to be sampled. To select a number to be dialed, a telephone prefix was first selected by a random process from among the prefixes serving the sampling area. Then by a second random process the suffix was chosen from among all possible suffixes to the selected prefix. This number, which may have been a business number, an unassigned number or even

an unlisted number as well as a listed residential number was called to select an individual for the study. Business numbers were not eligible for selecting controls. For non-answered calls to working numbers, repeat calls were made on a specified schedule which assured at least six calls; two on weekday mornings, two on weekday evenings and two on weekends, until either an answer was received or six calls were placed. For calls made to a residence, a census of all household members was taken and from among the eligible members (if any) one person was selected, also by a random process, and invited to participate in the study.

The possibility of bias by the RDD technique exists if different proportions of controls had multiple phones or had no phones through which they could be reached. Because all cases had telephones it was permissable to limit control selection to those who could also be reached by phone. A small proportion, approximately the same among cases and controls, had two or more telephones in their household.

Other sources of bias, not peculiar to this control selection technique but which are hazards to the general process of control selection in case-control studies, include a "response" bias. This bias can occur in several ways. It is possible that proportionally more (or fewer) cases than controls agree to participate in the study and that those who refuse to participate differ from the participants in some way that is related to one or more variables under study. In this study, another source of response bias was possible. This was the possibility of different response rates in different areas. In fact this was the case and will be discussed in a later section.

DATA REVIEW

Final Case-Control Counts

After key data entry, the process of editing the file began. Each variable was checked for completeness and interfield comparisons for consistancy and accuracy were performed. At the beginning of this process there were a total of 268 cases and proxies and a total of 410 controls. Twenty-six respondents were found to have incomplete data fields and were deleted. Fourteen of the 19 cases deleted due to incomplete data were proxies. In addition, 7 controls were deleted due to incomplete data.

There were a few age, race and sex combinations in which no cases occurred during the study period, even though one or two were anticipated. Controls collected for those particular combinations were dropped. Altogether, thirty controls were dropped for a lack of corresponding cases. At the end of the editing and matching processes 19 cases and 37 controls had been deleted leaving 249 cases and 373 controls for analysis. Through interviewing either proband or proxy cases, 75% of the eligibles were available for analysis. These results are summarized in Table 14.

TABLE 14. DISPOSITION OF THE 332 CASES ELIGIBLE FOR INTERVIEW AND ANALYSIS IN THE CONTRA COSTA COUNTY LUNG CANCER CASE-CONTROL STUDY

	Probands	Proxies
Eligibles	238	94
Refused participation	12	
Dead, no proxy available	11	
Ill	5	
Moved	11	
Other	25	
Incomplete questionnaire	5	14
Total deletions	69	14
Interviewed and analyzed	169	80

Generated Variables

Additional variables were generated from information collected in the questionnaire or in combination with other data concerning air pollution, residential water supply, or county population estimates.

Air Pollution

The measure of the respondent's exposure to air pollution was expressed as an estimated cumulative dose in Contra Costa County for each pollutant. In the air monitoring component of this study, census tracts in Contra Costa County were assigned air pollutant values based upon measurements taken at various monitoring stations throughout the county and constructed by interpolation of air pollution station measurements. Air pollutant values for each census tract were those values at each census tract population centroid. A total air pollutant dose for each study participant was calculated from the length of residence in each census tract. Thus, an air pollutant dose is the sum of the products of the duration of residence in each census tract times the corresponding air pollutant values for each census tract. Because the purpose of this analysis was to evaluate the effect of air pollution in Contra Costa County on lung cancer incidence, residence time outside the county was not given any value. This approach does not create any bias although recent immigrants to the county contribute little information to the analysis.

Smoking

The respondents' smoking experience was characterized by several parameters; total smoking duration, total pack years and average packs smoked per day. The respondents' smoking history was recorded by separate smoking periods. Each period was defined by the number of cigarettes (later converted to packs) smoked per day and the number of years of smoking at that level. Thus, total smoking duration is the sum of the durations for each smoking period. Pack years is the sum of the products of the number of packs smoked per day and the duration for each period. Average packs smoked per day is pack years divided by total smoking duration.

Occupation

The analysis for occupational exposure associations was based on the classification of occupation and industry titles in the 1980 Alphabetical Index of Industries and Occupations (II0), (Bureau of Census, 1980). For each work experience, an occupation and industry code was assigned.

The employment groups considered to be at risk of possible exposure to hazardous substances in the analysis were derived from the occupations listed under the II0 classifications of precision production, craft and repair occupations, operators, fabricators, and laborers. Four major industrial groups: petrochemical, metal, construction and all "other" remaining industries were also created.

"Construction" is a single industrial classification. "Petrochemical" industries include the manufacture of chemicals and allied products, petroleum and coal products, and rubber and miscellaneous plastics products. "Metal" industries include metal mining, primary metal industries such as blast furnaces and foundries, metal product manufacturing, and machinery manufacturing including electrical and transportation equipment.

The remaining group of industries, "Other", includes agriculture; the manufacturing categories of food and tobacco, textiles, paper and printing, leather, lumber and furniture, stone products, and professional equipment; the transportation, communication and public utilities industries; wholesale and retail trade; repair and personal services; and professional and public administration.

Over the entire employment history, if the worker fell into an indicated employment group and industry group, the duration of time worked in the industrial group was calculated and accumulated.

An asbestos exposure variable was created from various occupational categories. All shipyard occupations plus all other jobs for which asbestos exposures were reported by the respondent were combined to form a total duration of asbestos exposure per respondent. This variable is not mutually exclusive of other occupational categories. No attempt was made to quantify individual asbestos exposures other than as job duration.

Water Source

The distribution of the water sources in Contra Costa County were available from the Sanitary Engineering Branch of the Health Services Department. The three main water sources for Contra Costa County are Mokelumne River, San Pablo Reservoir, and Sacramento River Delta water systems. The census tract population centroid and the distribution of the water sources were used to assign a water source to every census tract. In turn, each participant was assigned a water source based on the census tract of residency at the time of interview or, for cases, diagnosis.

Toxic Waste Dumps

Certain of the census tracts in Contra Costa County contain known dumps of toxic or chemical waste. These census tracts were combined to form a toxic waste dump area for comparison to all other tracts.

Sample Adjustment Factor

In addition to matching the controls to the percent distribution of cases, they were also selected to represent the geographic distribution of the non-cancer population. To evaluate possible response variation, the number of controls expected from each census tract was computed and compared to the number actually obtained. It was found that one area of the county was overrepresented among controls and a separate small area of the county was underrepresented. Therefore the responses of the controls from those two areas were appropriately weighted in analyses to eliminate any effect of the response bias.

To assess the results of random sampling for controls within age, race, sex categories, the observed results were compared with expected numbers computed as follows.

For each age group (under 50, 50 and over), race (white, black), sex (male, female) category (k), the expected number of controls for a census tract was computed as the proportion of the population of category k residing in census tract "x" times the total number of sampled controls observed in the study for category k. Mathematically this can be expressed

$$E_{xk} = \frac{P_{xk}}{TP_k} * C_k$$

where, E_{xk} = expected number of controls for tract x, category k.

P_{xk} = estimated population of tract x, category k.

TP_k = total population in the county of category k.

C_k = total number of sampled controls of category k.

Summing across all (k) categories for a given tract yielded the total number of controls E_x expected for tract (x).

$$E_x = \sum_{k=1}^{8} E_{xk}$$

where, E_x = total number of expected controls for tract (x).

An over- or under-response rate of observed controls in the tract was said to exist if the statistical quantity, chi-square,

$$X^2 = \frac{(O-E)^2}{E} \quad \text{exceeded 3.5.}$$

The result of this comparison showed an over-response rate in 13 census tracts of the Richmond-San Pablo area of the county. The ratio of observed to expected in these tracts was as much as 9 times greater than expected. An under-response rate, as low as one-tenth of expected was found to have occurred in two census tracts in Concord, one in Antioch, one in the Brentwood area, and one in Pinole (see Figure 6).

An adjustment factor variable, the observed divided by expected, was therefore assigned for each control meeting the above criteria. This factor was incorporated in all statistical analyses to adjust for the variable response rate.

Alcohol Intake

Alcohol consumption was estimated on the basis that the amount of alcohol in an average glass of wine, a can of beer and a jigger of spirits or liquor is approximately the same. The number of each of these drinks consumed per week was added to form a total number of servings per week, treated as a continuous variable.

Matching Within Age, Race, and Sex Strata

For the purpose of subsequent statistical analysis, i.e., multiple logistics regression with matched case-control data, it was necessary to stratify the case-control population such that each case was matched to controls with identical race, identical sex, and approximately equal age. A variable matching ratio was used.

TABLE 15. RESULTS OF MATCHING CONTROLS TO CASES BY SEX AND MATCHING RATIO

The number of (249) cases matched to the indicated number of (373) controls

Sex	Matching ratio 1	2	3	4	5	Total
Male	80	58	4	1	1	144
Female	65	29	11	0	0	105
Total	145	87	15	1	1	249

Within a particular age, race, sex stratum an algorithm for matching cases to controls by age was developed. Within a stratum cases and controls were separately ranked from low to high by ascending age. The number of controls assigned to a case was given by the following formula:

$$\text{No. of controls for case (i)} = \text{INTEGER}(.5 + CR_i * MR) - \text{INTEGER}(.5 + CR_{i-1} * MR)$$

where, CR_i = case rank number,

MR = matching ratio = (No. of controls in cell)/(No. of cases in cell)

INTEGER = use the number to the left of the decimal result of the computation within parentheses.

For the first case, $CR_{i-1} = 0$.

The algorithm produced a variable matching ratio in which a case may have one or more matched controls. The results of this process are summarized in Table 15.

DATA ANALYSIS

The Method of Multiple Logistics

Multivariate risk analysis is motivated by the need for methods that assess the direct predictive strength associated with each member of a cluster of interrelated risk factors. With univariate analysis, each measure is examined separately for its relation to disease incidence. However, only multivariate risk analysis provides a method for determining whether (1) both measures have direct predictability for disease, or (2) one appears as a predictor in the univariate sense wholly or in part because of its association with the other.

The essence of multiple risk analysis is to examine changes of risk with varying levels of one factor at various fixed levels of the remaining factors. When many risk factors are being considered, a traditional method for multiple risk analysis based on cross-tabulation is not feasible. Erratic results occur because too few cases of disease must be distributed over too many risk subgroups. Instead, multiple risk analysis can be approached with use of risk models.

The multiple logistic risk model is the most widely used. It is constructed so that risk of disease is automatically restricted to the appropriate range, 0 to 1. In this model risk, R, is represented by the equation

$$R = 1/\left[1 + e^{-(B_0 + B_1X_1 \ldots + B_kX_k)}\right]$$

where B_0, B_1, $\ldots B_k$ are the k + 1 logistic coefficients. These are estimated from data which measure levels for k risk factors $(X_1, X_2, \ldots, X_k)$ for each subject.

The rather formidable logistic model can be simplified somewhat by linearizing it to give $\ln R/(1 - R) = B_0 + B_1X_1 + B_2X_2 \ldots + B_kX_k$. The quantity, $\ln R/(1 - R)$, is called the "logit" of risk and is the mathematical transformation of risk from which this model derives its name. By thinking in terms of logit of risk, which ranges from minus infinity to plus infinity as risk ranges from 0 to 1, it is possible to interpret the logistic coefficients, $B_1, B_2, \ldots, B_k$. They measure the change in logit of risk per unit change in the respective risk factors.

Descriptive Statistics

Tables 16 and 17 present the mean and standard error for a number of variables used in the analysis. Data are given by sex and case-control status.

TABLE 16. DESCRIPTIVE STATISTICS FOR MALES. MEANS AND STANDARD DEVIATIONS FOR CASES (N=144) AND CONTROLS (N=217)

	Cases		Controls	
Variable Name	Mean	Standard deviation	Mean	Standard deviation
Air pollutants				
Total suspended particulate*	1424.30	942.55	1275.88	803.58
Lead*	11.49	7.58	10.80	6.20
Sulfate*	176.01	114.33	160.65	99.76
Nitrate*	146.42	94.49	136.24	77.46
Benzene soluble organics*	95.72	63.59	88.75	50.85
Mutagenicity**	122.96	77.95	119.83	75.27
Mutagenicity with S9**	149.63	94.47	146.03	89.18
Benzo(a)Pyrene***	3.78	3.64	3.53	2.56
Benz(a)Anthracene***	2.57	2.78	2.29	1.81
Chrysene***	3.12	3.12	2.80	2.00
Benzo(ghi)Perylene***	22.65	18.00	22.05	14.18
Industry, duration (yrs)				
Asbestos related	3.50	7.79	1.91	5.70
Petrochemical	1.57	5.62	1.35	5.60
Construction	2.63	8.10	2.58	6.09
Metal	3.38	8.09	1.88	6.53
Other	12.08	13.78	7.42	11.57

(continued)

TABLE 16 (continued)

Variable Name	Cases Mean	Cases Standard deviation	Controls Mean	Controls Standard deviation
Water source (yes/no)				
Mokelumne River	37%	4%	52%	3%
San Pablo	21%	3%	13%	2%
Sacramento River Delta	42%	4%	35%	3%
Lifestyle				
Smoker (ever/never)	97%	1%	74%	3%
Smoking duration (yrs)	38.10	13.45	24.95	18.58
Smoking (pack yrs)	53.97	35.15	26.19	27.86
Smoking				
(Average # packs/day)	1.34	0.70	0.79	0.70
Alcohol (servings/week)	20.95	34.07	10.61	16.21
Diet				
Green vegetables (servings/week)	5.17	2.42	5.59	2.30
Yellow vegetables (servings/week)	4.28	2.49	4.44	2.46
Demographic				
Education (yrs)	12.80	6.00	14.00	3.20
Income****	$17,000	$8,000	$21,500	$8,200

*Yr x $\mu g/m^3$

**Yr x rev/m^3

***Yr x ng/m^3

****Cases (N = 113), Controls (N = 184)

TABLE 17. DESCRIPTIVE STATISTICS FOR FEMALES. MEANS AND STANDARD DEVIATIONS FOR CASES (N = 105) AND CONTROLS (N = 156)

	Cases		Controls	
Variable name	Mean	Standard deviation	Mean	Standard deviation
Air pollutants				
Total suspended particulate*	1155.34	879.59	1284.54	977.82
Lead*	9.59	7.33	10.96	8.31
Sulfate*	144.85	113.31	167.38	130.39
Nitrate*	120.93	88.92	129.49	94.73
Benzene soluble organics*	79.86	59.98	86.72	64.06
Mutagenicity**	107.27	79.33	112.41	83.82
Mutagenicity with S9**	130.66	96.33	104.75	104.81
Benzo(a)Pyrene***	3.19	3.01	2.89	2.84
Benz(a)Anthracene***	2.17	2.27	1.85	2.04
Chrysene***	2.52	2.27	2.52	2.63
Benzo(ghi)Perylene***	19.31	15.79	18.87	15.50
Industry, duration (yrs)				
Asbestos related	0.32	1.67	0.18	0.59
Petrochemical	0.17	1.3	0.02	0.20
Construction	0.32	3.30	0.01	0.14
Metal	0.54	3.11	0.32	1.81
Other	1.39	3.81	1.69	5.62

(continued)

TABLE 17 (continued)

Variable name	Cases Mean	Cases Standard deviation	Controls Mean	Controls Standard deviation
Water source (yes/no)				
Mokelumne River	45%	5%	43%	4%
San Pablo	17%	4%	21%	3%
Sacramento River Delta	38%	5%	36%	4%
Lifestyle				
Smoker (ever/never)	86%	3%	58%	4%
Smoking duration (yrs)	33.11	15.91	19.98	19.27
Smoking (pack yrs)	39.30	27.29	14.27	17.52
Smoking				
(Average # packs/day)	1.03	0.66	0.43	0.48
Alcohol (servings/week)	10.93	23.61	5.67	9.20
Diet				
Green vegetables (servings/week)	5.57	2.71	5.97	2.67
Yellow vegetables (servings/week)	4.22	2.21	4.27	2.62
Demographic				
Education (yrs)	12.76	.96	13.92	2.16
Income****	$13,700	$8,700	$14,900	$7,900

*Yr x $\mu g/m^3$

**Yr x rev/m^3

***Yr x ng/m^3

****Cases (N = 79), Controls (N = 120)

Analytic models

Fitting Smoking Variables

Smoking is known to be the single largest cause of lung cancer in the population. As shown in Table 16 of this report, 97% of male cases fit the criterion for being a regular smoker as compared to 74% of the controls. The comparable figures for females (Table 17) are 86% of cases and 58% of controls. Consequently, it was important to identify the best measurement representing smoking so that all other variables could be examined, adjusted for the effect of smoking. The "best" smoking measurement was that smoking variable or combination of smoking variables which explained the largest amount of lung cancer.

To account adequately for smoking and its relationship to lung cancer in this data set, a number of logistic models of lung cancer and smoking were tested and reviewed.

Available smoking variables were 1) ever or never smoked regularly, i.e., one or more cigarette per day for 12 months, 2) duration, i.e., total number of years smoked, 3) average packs per day smoked, and 4) total pack years, i.e., packs per day times duration smoked. Variables were "fitted" singly and in pairs for males and females as separate analytic groups. Interaction terms, as well as various transformations (e.g., the square of the variables) were modeled.

A simple "best fit" model was selected. Two measures, smoking duration and average packs per day used as independent terms in a single model, proved to be statistically significant for males. For females, only the average packs per day proved to be statistically significant, but for the purpose of parallel presentation, the smoking duration variable was also retained. Therefore, both measures of smoking (duration and average packs per day) were used in all subsequent analytical models.

Analysis with Additional Variables

Table 18 presents the odds ratios computed via the multiple logistics analysis technique for two environmental factors, both with and without controlling for the effect of smoking. None of the odds ratios are statistically significant.

Table 19 presents the odds ratios for several of the measured air pollutants in Contra Costa County, controlled for known risk factors. The dose values pertain to the total cumulated exposure from residence in the county. Four examples for males are presented. No value for any pollutant for either males or females approached statistical significance, with or without controlling for known risk factors.

TABLE 18. RISK* OF LUNG CANCER, BY SEX, ASSOCIATED WITH SELECTED ENVIRONMENTAL FACTORS IN CONTRA COSTA COUNTY, WITH AND WITHOUT CONTROLLING FOR SMOKING

	Males		Females	
Factor	OR	p**	OR	p**
Industrial area				
Not controlled	1.68	0.06	0.71	0.40
Controlled for smoking	1.55	0.17	0.53	0.21
Near waste dumps				
Not controlled	1.04	0.94	0.76	0.75
Controlled for smoking	0.97	0.95	0.67	0.67

*Risk expressed as the odds ratio (OR)

**Significance probability

TABLE 19. RISK* OF LUNG CANCER FOR VARIOUS AIR POLLUTANTS, CONTROLLED FOR SMOKING, DRINKING AND ASBESTOS EXPOSURE, CONTRA COSTA COUNTY MALES

Factor	OR	p**
SO_4 dose ($\mu g/m^3$ from county residence)	1.002	.155
NO_3 dose ($\mu g/m^3$ from county residence)	1.003	.142
Mutagen dose (rev/m^3 from county residence)	1.001	.460
Benzo(a)pyrene (ng/m^3 from county residence)	1.032	.505

*Risk expressed as the odds ratio (OR)

**Significance probability

TABLE 20. RISK* OF LUNG CANCER FOR VARIOUS OCCUPATIONAL CATEGORIES, CONTROLLED FOR SMOKING, DRINKING, ASBESTOS EXPOSURE AND SO_4 DOSE, CONTRA COSTA COUNTY MALES

Factors	OR	p**
Metal industry (per year of employment)	1.016	432
"Other" industries (per year of employment)	1.009	.430
Construction industry (per year of employment)	1.024	.254
Petrochemical industry (per year of employment)	0.991	.713

*Risk expressed as the odds ratio (OR)

**Significance probability

Table 20 presents the odds ratios for the broad occupational categories of male blue collar worker. None of the odds ratios approached statistical significance, with or without controlling for known risk factors.

Tables 21 and 22 present the odds ratios for each factor included in the saturated model for males and females, respectively. This multiple logistic analysis using 13 variables did not substantially alter the odds ratio for any factor from that found in simpler models. Further, this saturated model failed to explain much more of the lung cancer risk than did the simpler models using only variables with either a statistically significant relationship to lung cancer or a known causal relationship to lung cancer established from other studies.

TABLE 21. RISK* OF LUNG CANCER FOR EACH FACTOR IN THE SATURATED MODEL FOR MULTIPLE LOGISTIC ANALYSIS, CONTRA COSTA COUNTY MALES

Factor	Unit	OR	p**
Smoking duration	(per year smoked)	1.039	< .0005
Smoking dose	(per pack per day)	1.896	.007
Green vegtables	(per weekly serving)	0.809	.012
Alcoholic drinks	(per weekly serving)	1.008	.237
Asbestos exposure	(per year of exposure)	1.035	.188
SO_4 dose	(total $\mu g/m^3$ from county)	1.001	.547
Petrochemical industry	(per year of employment)	0.987	.617
Construction industry	(per year of employment)	1.025	.293
Metal industry	(per year of employment)	1.021	.354
"Other" industries	(per year of employment)	1.008	.541
Mokelumne River water	(source at interview)	0.620	.280
Sacramento delta water	(source at interview)	0.724	.488
Yellow vegetables	(per weekly serving)	1.022	.780

*Risk expressed as the odds ratio (OR).

**Significance probability

TABLE 22. RISK* OF LUNG CANCER FOR EACH FACTOR IN THE SATURATED MODEL FOR MULTIPLE LOGISTIC ANALYSIS, CONTRA COSTA COUNTY FEMALES

Factor	Unit	OR	p**
Smoking dose	(per pack per day)	10.249	.004
Smoking duration	(per year smoked)	0.998	.910
Green vegetables	(per weekly serving)	1.194	.176
Alcoholic drinks	(per weekly serving)	0.967	.252
Metal industry	(per year of employment)	1.210	.244
"Other" industries	(per year of employment)	0.928	.357
Yellow vegetables	(per weekly serving)	0.897	.442
Petrochemical industry	(per year of employment)	1.327	.477
Sacramento delta water	(source at interview)	1.072	.923
Mokelumne River water	(source at interview)	0.919	.906
Asbestos exposure	(per year of exposure)	1.069	.780
Construction industry	(per year of employment)	1.160	.917
SO_4 dose	(total $\mu g/m^3$ from county)	1.001	.811

*Risk expressed as the odds ratio (OR).

**Significance probability

TABLE 23. THE RISK* OF LUNG CANCER FOR MALES SMOKING THE AVERAGE DOSE AND DURATION OF SMOKING MALE CASES (N=140) AND FOR FEMALES SMOKING THE AVERAGE DOSE OF SMOKING FEMALE CASES (N=90)

	Males	Females
Average years smoked	39.2	38.5
Risk per year	1.04	NS**
Average packs smoked per day	1.38	1.21
Risk per pack	1.90	10.25
Total average smoking risk	11.28	15.47

*Risk expressed as the odds ratio

**Not used in the computation since there was no statistically significant effect

To assess the approximate contribution to lung cancer risk by smoking, several methods were employed. The first estimated the effect from the average daily dose and the average smoking duration among all male cases who were smokers and for female cases, the average daily dose smoked among smokers. The results, presented in Table 23, illustrate that the average effect of the statistically significant smoking variables on the risk for lung cancer is similar for both males and females, e.g., 11.28 and 15.47 times that of nonsmokers, respectively.

The second method computed the proportion of the "explainable" lung cancer contributed by smoking, under alternate conditions (Tables 24 and 25) (Coles, 1980). There are three alternative conditions for accounting for lung cancer by the variables under study. The first is to require that only those factors statistically significantly related to lung cancer in the analysis be considered candidates for a causal relationship. This is the most conservative condition and, under this condition, the amount of lung cancer "explained" by the data, and the analytic model, is only that amount "explained" by those factors bearing a statistically significant relationship to lung cancer. Under this conservative condition, 29% of all lung cancer among females and 27% of all lung cancer in males can be accounted for by this analysis. Of that part which can be accounted for, 100% in females and 83% in males is contributed by smoking.

The second alternative condition is to allow those factors with statistically significant relationships to lung cancer plus those factors that have been shown by other studies to have a causal relationship to lung cancer (even if in this study the relationship does not reach statistical

TABLE 24. THE RELATIVE CONTRIBUTION BY EACH RISK FACTOR, UNDER THREE ASSUMPTIONS*, TO THE EXPLAINABLE PROPORTION OF MALE LUNG CANCER IN CONTRA COSTA COUNTY, EXPRESSED AS CUMULATIVE PERCENTS

Risk factor	p	Cumulative percent contributed by Assumption* 1	2	3
Smoking dose	.007	83.3	76.7	70.9
Smoking duration	<.0005			
Green vegetables	.012	100.0	92.1	85.1
Alcoholic drinks	.237	-	95.8	88.5
Asbestos exposure	.188	-	100.0	92.3
SO^4 dose	.547	-	-	93.8
Petrochemical industry	.617	-	-	94.1
Construction industry	.293	-	-	95.8
Mokelumne River	.280	-	-	100.0
Sacramento Delta	.488			
Metal industry	.354			
Other industries	.541			
Yellow vegetables	.780			
Proportion of all lung cancer explained:		.268	.291	.315

*Assumption 1. The most conservative assumption: Lung cancer can be causally related only to those factors to which it is statistically significantly related ($p<0.05$).

Assumption 2. The moderate assumption: Lung cancer may be causally related to those factors to which it is statistically significantly related ($p<0.05$) and also to those factors known from other information to be causally related.

Assumption 3. The most liberal assumption: Lung cancer may be causally related to all factors studied, irrespective of the statistical significance of the relationship.

TABLE 25. THE RELATIVE CONTRIBUTION BY EACH RISK FACTOR, UNDER THREE ASSUMPTIONS*, TO THE EXPLAINABLE PROPORTION OF FEMALE LUNG CANCER IN CONTRA COSTA COUNTY, EXPRESSED AS CUMULATIVE PERCENTS

Risk Factor	p	Cumulative Percent Contributed by Assumption* 1	2	3
Smoking dose	.004	100.0	93.2	81.1
Smoking duration	.910	-		
Green vegetables	.176	-	95.5	83.1
Alcoholic drinks	.252	-	99.8	86.9
Asbestos exposure	.780	-	100.0	87.0
SO_4 dose	.811	-	-	87.0
Petrochemical industry	.477	-	-	90.5
Construction industry	.917	-	-	91.0
Mokelumne River	.906			
Sacramento Delta	.923			
Metal industry	.244	-	-	100.0
Other industries	.357			
Yellow vegetables	.442			
Proportion of all lung cancer explained:		.294	.316	.363

*Assumption 1. The most conservative assumption: Lung cancer can be causally related only to those factors to which it is statistically significantly related ($p<0.05$).

Assumption 2. The moderate assumption: Lung cancer may be causally related to those factors to which it is statistically significantly related ($p<0.05$) and also to those factors known from other information to be causally related.

Assumption 3. The most liberal assumption: Lung cancer may be causally related to all factors studied, irrespective of the statistical significance of the relationship.

significance) to be considered as candidates for causal relationships. This is a moderate condition and allows the effect of alcohol use and asbestos exposure to help account for lung cancer, even though in this analysis neither were found to bear a statistically significant relationship to lung cancer. Given that asbestos and alcohol consumption are generally accepted risk factors for lung cancer, the lack of statistical significance may be due to either a very small contribution by each factor to lung cancer in this study or that the cases and controls did not differ much in their asbestos and alcohol exposure. Under this moderate condition, 32% of the lung cancer among females and 29% of the lung cancer among males can be "explained". Of the explainable portion, 93% for females and 77% for males is contributed by smoking.

The third and most liberal condition is to allow all variables to be considered eligible candidates for accounting for lung cancer, regardless of the statistical significance of their relationship to lung cancer. Under this condition a maximum of 36% among females and 32% of the lung cancer among males is explainable by the saturated analysis model. Of this explainable amount, smoking contributes 81% among females and 71% among males.

While this last condition is clearly unacceptable for the purposes of concluding causal relationships, it has the benefit of demonstrating the maximum contribution by those variables studied if they were to have a causal role. For example, even though SO_4 exposure does not have a statistically significant relationship to lung cancer in this analysis, it is possible to estimate that only about 1.5% of the total explainable male lung cancer (i.e., less than 1/2% of all male lung cancer) and none of the female lung cancer could be attributed to SO_4 exposure if it were a causal factor.

DISCUSSION

In any analysis a major goal is to be able to explain as much of the disease of interest as possible with the information gathered for analysis by the study. In this study several variations on this goal were also sought. One was to determine whether any lung cancer could be attributed to constituents of air pollution. The other was to explain the simultaneous occurrence of three factors in Contra Costa County: higher incidence of lung cancer in males, higher proportions of blue collar workers and higher levels of SO_4 in ambient air. Therefore the analyses also sought to explain the earlier findings and to reconcile as many of the facts regarding lung cancer in Contra Costa County as possible.

It is unrealistic to expect that a study explain, through its gathered data, all of the cases of a disease with multiple causation. Reasonable expectations are that observed differences in the rate of disease between areas or groups be explained and that the information gathered account for a significant portion of the disease.

TABLE 26. COMPARISON OF THE RISK* OF LUNG CANCER, BY SEX, FOR AMBIENT AIR SULFATE EXPOSURE MEASURED AS THE AVERAGE ANNUAL LEVEL OF THE CENSUS TRACT RESIDENCE AT DIAGNOSIS (SO_4 LEVEL) AND AS THE COMPUTED TOTAL LIFETIME OF DOSE IN CONTRA COSTA COUNTY (SO_4 DOSE)

Factor	Males		Females	
	OR	p**	OR	p**
SO_4 level	1.26	0.01	0.94	0.78
SO_4 dose	1.00	0.23	1.00	0.41

*Risk expressed as the odds ratio (OR)

**Significance probability

To determine whether the data collected at the individual level corroborated the previous findings from this project, several analyses were conducted that did not neccesarily test causal hypotheses. The earlier incidence analysis showed that residents of the Industrial area experienced lung cancer at a rate approximately 40% higher than residents of the Non-Industrial area for the period 1975-79. The cases and controls in this component of the study were categorized as to their residential address at the time of interview (for controls) or at the time of initial diagnosis (for cases). Table 18 shows that when male participants are so categorized there is an elevated odds ratio ($p = 0.06$) associated with residence in the Industrial area, as previously found. When smoking is taken into account, however this relationship is essentially destroyed ($p = 0.17$), indicating that in this set of data, the difference in lung cancer risk for males between Industrial and Non-industrial areas can be substantially accounted for by a difference in smoking practices of the males of the two areas. For females no elevated odds ratio for residence in the Industrial area exists.

Proximity to known locations of toxic waste dumps had no effect on lung cancer risk, as shown in Table 18.

The previous correlation analysis produced an association between ambient SO_4, percent blue collar workers and lung cancer among males. That analysis used, as a measure of SO_4, the level assigned to the census tract at the time of diagnosis. Table 26 shows that for males, a positive but not statistically significant relationship exists between the level of SO_4 in the census tract at the time of diagnosis (for cases) or interview (for controls). If ambient SO_4 exposure really imparts an increased risk of lung cancer one would expect a more precise measure, estimated cumulative dose of ambient SO_4 in the county, to demonstrate a stronger relationship. However, when the total cumulated dose of ambient SO_4 in Contra Costa County is considered, no relationship exists between SO_4 and lung cancer. Adjusting for smoking or other factors, whether significant or not,

did not materially change this finding.

This analysis examined the effect of smoking and air pollution in considerable detail. No measure of air pollution was found to have a statistically significant relationship to lung cancer. In considering these negative findings, the following must be weighed.

1. The estimated cumulative dose of each pollutant was computed from values in Contra Costa County only. If the population in Contra Costa County were highly mobile it could dilute the contribution by each respondent such that a weak relationship could not be detected. In this situation, however, no correlation with the pollutant level at the place of residence at diagnosis would represent a causal relationship unless the relationship were both immediate and transitory, a requirement not met by any other known carcinogenic agents.

2. Air pollutants were measured for only one year. Thus if that year were atypical for the relative levels of each pollutant from area to area or if measurement of levels were not typical of the census tract, a real relationship would be obscured through misclassification.

3. Another assumption is that each subject is locationally fixed at his/her census tract of residence; consequently this excludes any other locational exposure such as those in the workplace.

4. Only solid air pollutants were measured. At the start of the project no reliable method existed for measuring the mutagenicity of volatile gases. Also, the project was not funded to measure levels of volatile gases, such as benzene, which could have biologic effects at very low doses. However, in general the volatile and particulate pollutants originate from the same emission sources and the general air flow patterns remain the same from year to year so that if there were a significant carcinogenic effect from a volatile pollutant that was not measured, it would be expected to exhibit its effect in a distribution similar to that of a particulate pollutant from the same source.

Even with these caveats, the failure to detect any effect on lung cancer by air pollution suggests that no significant effect exists.

The proportion of lung cancer explainable by this analysis was only about a third, most of which was contributed by smoking. This may be due to at least three reasons. The first is that the multiple logistics analysis draws its information about risk from differences between cases and controls. With regards to smoking, a high proportion of both cases and controls smoked. This may have led to an underestimate of the risk from smoking.

The occupational categories are very broad and undoubtedly contain specific occupations that are of higher and lower risk than the mixture that

constitutes the broad category. The occupational analysis therefore likely explains less lung cancer than potentially it could. This would be expected to be more true of males than of females since in general, males would be expected to have a higher proportion of their numbers in occupations with carcinogenic hazards. This supposition is supported by the fact that higher proportion of lung cancer among females is explained, under any assumption, than among males.

Lastly, a major determinant for any cancer was not measurable by this study. That factor is individual susceptibility. It may explain a significant proportion of lung cancer but is not identifiable by a comparison of exposures of cases and controls.

SUMMARY AND CONCLUSIONS

This analysis of case-control data suggests that the major contributor to lung cancer in Contra Costa County is smoking. Further, smoking accounts for most of the previously identified difference in lung cancer incidence between the Industrial and Non-industrial areas. Because of the high prevalence among both cases and controls, the contribution of smoking may be underestimated.

There was no identified effect on lung cancer risk contributed from any measured constituent of air pollution. The one air pollutant (SO_4) significantly correlated with male lung cancer incidence in the indirect correlational analysis, in this case-control analysis had a positive but not statistically significant relationship with lung cancer risk only when SO_4 level at the current address was used as the measurement. When a measure of total lifetime dose of SO_4 from Contra Costa County was used, no elevated risk was apparent.

One dietary factor had a significant ($p = 0.01$) protective effect for males and a similar but not statistically significant ($p = 0.18$) effect for females. This factor, weekly servings of green vegetables, is a crude measure for several dietary constituents believed to reduce the risk of cancer of several types. Both vitamin A and cruciferous vegetables would be included in this dietary measure. The dietary measure, weekly servings of yellow vegetables, did not discriminate between cases and controls.

None of the broad occupational categories had any significant relationship to lung cancer risk in males. A more detailed analysis of the effect of various occupations on lung cancer risk is desirable and support for this subsequent analysis is being sought.

The effects of alcohol and asbestos exposure, as measured, did not bear a statistically significant relationship to lung cancer in this analysis. In any subsequent analysis a more quantitative measure of asbestos exposure would be desirable.

There was no apparent effect of source of drinking water or proximity to known toxic waste dumps on the risk of lung cancer.

These data confirm the known causal relationship between smoking and lung cancer. They provide some reassurance that constituents of particulate air pollution do not contribute measurably to the risk of lung cancer. This is consistent with the findings of several other studies. These data provide supportive evidence for the protective effect of dietary factors on cancer risk, a finding consistent with other epidemiologic and laboratory studies.

BIBLIOGRAPHY

AIHL. Determination of Total Organic Materials in Atmospheric Particulate Matter. AIHL Method No. 65. California State Department of Health Services, Berkeley, California, 1975.

AIHL. The Chemical and Biochemical Characterization of Particulate Matter as Part of an Epidemiological Cancer Study. AIHL Method No. 61. California State Department of Health Services, Berkeley, California, 1980a.

AIHL. Determination of Nitrates in Atmospheric Particulate Matter (Brucine Method). AIHL Method No. 66. California State Department of Health Services, Berkeley, California, 1980b.

AIHL. Determination of Sulfate in High Volume Particulate Samples: Turbidimetric Barium Sulfate Method. AIHL Method No. 61. California State Department of Health Services, Berkeley, California, 1980c.

Ames, B., J. McCann and E. Yamasaki. Method for Detecting Carcinogens and Mutagens with the Salmonella/Mammalianmicrosome Mutagenicity Test. Mutation Res., 31:347-364, 1975.

BAAQMD. Total Suspended Particulate Gravimetric Analysis Procedure. Bay Area Air Quality Management District, San Francisco, California, 1977.

Beamis, J. F., A. Stein and J. L. Andrews, Jr. Changing Epidemiology of Lung Cancer. Increasing Incidence in Women. Med. Clin. N. Am., 59(2):315-324, 1975.

Blot, W. J., L. A. Brinton, J. F. Fraumeni, Jr. and B. T. Stone. Cancer Mortality in U.S. Counties with Petroleum Industries. Science, 198, 51-53, 1977.

Blot, W. J., J. F. Fraumeni, Jr. Arsenical Air Pollution and Lung Cancer, The Lancet, July 26, 1975.

Buell, P. and J. E. Dunn, Jr. Relative Impact of Smoking and Air Pollution on Lung Cancer. Arch. Env. Hlth. 15:29'-29?, 1967.

Burbank, F. U.S. Lung Cancer Death Rates Begin to Rise Proportionately More Rapidly for Females Than for Males: A Dose Response Effect? J. Chron. Dis. 25:473-479, 1972.

Bureau of the Census. 1980 Alphabetical Index of Industries and Occupations. U.S. Department of Commerce, Washington, D.C., 1981. 267 pp.

Carnow, B. W. The "Urban Factor" and Lung Cancer: Cigarette Smoking or Air Pollution? Environmental Health Perspectives. 22:17-21, 1978.

Clemmesen, J. Registration of Data Concerning Cancers Possibly Produced by Air Pollution, In: Air Pollution and Cancer in Men, U. Mohr, and T. L. Schmahl, eds. IARC Scientific Publ. No. 16. International Agency for Research on Cancer, Lyon, 1977.

Coles, et al. Determining the Most Valuable Clinical Models. Methods Information In Medicine, 19(1):41-49, 1980.

del Regato, J. A. and H. J. Spjut. Ackerman and del Regato's Cancer Diagnosis and Prognosis. The C.V. Mosby Company, St. Louis, Missouri, 1977.

Doll, R. and R. Peto. The Causes of Cancer: Quantitative Estimates of Avoidable Risks of Cancer in the United States Today. JNCI 66(6):1191-1308, 1981.

Flessel, C., J. J. Wesolowski, et al. Integration of the Ames Bioassay and Chemical Analyses in an Epidemiological Cancer Incidence Study. pp. 67-83. In: Application of Short-Term Biossays in the Fractionation and Analysis of Complex Enviromental Mixtures, M. D. Water, S. Sandhu, et al, eds. Vol. 2. Plewum Publishing Corporation, New York, 1981.

Fraumeni Jr., J. F. Persons at High Risk of Cancer. An approach to Cancer Etiology and Control. Academic Press, New York, 1975.

Friberg, L. and R. Cederlof. Late Effects of Air Pollution with Special Reference to Lung Cancer. Environmental Health Perspectives, 22:45-66, 1978.

Gordon, R. J. Personal Communication, 1978.

Hammond, E. C. Smoking Habits and Air Pollution in Relation to Lung Cancer. In: Environmental Factors in Respiratory Disease, D. H. K. Lee, ed. Academic Press, New York, 1972.

Hammond, E. C. Tobacco, In: Persons at High Risk of Cancer. An Approach to Cancer Etiology and Control. J. F. Fraumeni, Jr., ed. Academic Press, New York, 1975.

Harvard University Laboratory for Computer Graphics and Spatial Analysis. SYMAP. Version 5.2, Harvard University, Cambridge, Massachusetts, 1975.

Henderson, B. E., R. J. Gordon, H. R. Menck, et al. Lung Cancer and Air Pollution in South Central Los Angeles County. Am. J. Epid. 101(6):477-488, 1975.

Higgins, I. T. Epidemiological Evidence on the Carcinogenic risk of Air Pollution, In: Air Pollution and Cancer in Man. U. Mohr, T. L. Schmahl, eds. IARC Scientific Publ. No. 16. International Agency for Research on Cancer, Lyon, 1977.

Linden, G. and D. F. Austin. A Rapid Reporting Cancer Incidence System. American Journal of Epidemiology. 99(3). The John Hopkins University, 1974.

Lombard, H. L. and E. P. Huyck. An Epidemiological Study of Lung Cancer Among Females. Growth, 32:41-56, 1968.

Moore, H. Application of Wavelength Dispersive X-Ray Fluorescence Spectrometry to the Determination of Lead in Atmospheric Particulate Matter Collected on High-Volume Glass Fiber Filters. AIHL Report 183. Air and Industrial Hygiene Laboratory, California Department of Health Services, Berkeley, California, 1976.

NCI. Third National Cancer Survey: Incidence Data. National Cancer Institute Monograph 41. Publication No. (NIH) 75-787, NCI, NIH, Public Health Service, U.S. Department of Health, Education and Welfare.

NCI. Incidence and Mortality Data, 1973-1977. National Cancer Institute Monograph 57. Publication No. (NIH) 81-2330. NCI, NIH, Public Health Service, U.S. Department of Health, Education and Welfare, 1981.

Oliphant, T. H. and R. B. McHugh. Least Significant Relative Risk Determination in the Case of Unequal Sample Sizes. American Journal of Epidemiolgy. 113(6). The John Hopkins University, 1981.

Pike, M. C., R. J. Gordon, et al. Air Pollution, In: Persons at High Risk of Cancer. An Approach to Cancer Etiology and Control. J. F. Fraumeni, Jr., ed. Academic Press, New York, 1975.

Pike, M. C., J. S. Jung, et al. In: Energy and Health. N. E. Breslow and A. S. Whittemore, eds. pp. 3-16. SIAM Institute for Mathematics and Society, Philadelphia, 1979.

Vena, J. E. Air Polution as a Risk Factor in Lung Cancer. Am. J. Epid. 116(1):42-56, 1982.

Williams, R. R. and J. W. Horn. Association of Cancer Sites with Tobacco and Alcohol Consumption and Socioeconomic Status of Patients: Interview Study from the Third National Cancer Survey. J. Natl. Cancer Inst. 58(3):525-547, 1977.

World Health Organization. International Classification of Diseases for Oncology. World Health Organization, Geneva, Switzerland, 1976.

Young, Jr., J. F., C. H. Percy, A. J. Asire, eds. Surveillance Epidemiology and End Results: Incidence and Mortality Data. National Cancer Institute Monograph 57. NIH Publ. 81-2330, National Cancer Institute, Bethesda, Maryland, 1981.