Environmental Health Perspectives
Vol. 54, pp. 319–332, 1984

Statistical Association between Cancer Incidence and Major-Cause Mortality, and Estimated Residential Exposure To Air Emissions from Petroleum and Chemical Plants

by John Kaldor,* John A. Harris,† Eva Glazer,‡ Sally Glaser,** Raymond Neutra,† Robert Mayberry,** Verne Nelson,‡ Lewis Robinson§ and Dwayne Reed††

An ecologic study design was used to investigate the relationship between exposure to air emissions produced by the petroleum and chemical industries, and average annual cancer incidence and major cause mortality rates among whites in Contra Costa County, California. Estimates for the exposure to major industrial sources of sulfur dioxide, hydrocarbons and oxides of nitrogen were used to subdivide the county by level of exposure to petroleum refinery and chemical plant emissions. Cancer incidence and major cause mortality rates were then calculated for whites in each of the exposure areas.

In both males and females, residential exposure to petroleum and chemical air emissions was associated with an increased incidence of cancer of the buccal cavity and pharynx. In males, age-adjusted incidence rates for cancers of the stomach, lung, prostate and kidney and urinary organs were also associated with petroleum and chemical plant air emission exposures. In both sexes, we found a strong positive association between degree of residential exposure and death rates from cardiovascular disease and cancer, and a less strong positive association between exposure and death rates from cerebrovascular disease. There was also a positive association in men for deaths from cirrhosis of the liver.

Although these observed associations occurred across areas of similar socioeconomic and broad occupational class, confounding variables and the "ecologic fallacy" must be considered as possible explanations. In particular, the stronger findings in men suggest an occupational explanation of the cancer incidence trends, and the effect observed in cirrhosis mortality suggests that lifestyle variables such as alcohol consumption were not adequately controlled for. While the public health implications of our findings remain unclear, the evidence presented is sufficient to warrant follow-up studies based on individual data in which possible biases can be more readily controlled.

Introduction

A large number of chemicals involved in petroleum refining and chemical manufacturing are toxic in various ways. In particular, many aromatic amines, nitrosamines, organometallics,

*Lawrence Berkeley Laboratory, Division of Biology and Medicine, Building 934, Berkeley, CA 94720. Author to whom correspondence should be addressed. Present address: Unit of Biostatistics and Field Studies, International Agency for Research on Cancer, 150, cours Albert-Thomas, F-69372 Lyon Cedex 08 France.

†Epidemiological Studies Section, California State Department of Health Services, 2151 Berkeley Way, Room 515, Berkeley, CA 94704.

‡Resource for Cancer Epidemiology, 5850 Shellmound, Emeryville, CA 94608.

**Program in Epidemiology, University of California, Berkeley, CA 94720.

§Bay Area Air Quality Management District, 939 Ellis Street, San Francisco, CA 94109.

††Honolulu Heart Program, 347 North Kuakini Street, Honolulu, Hawaii 96817.

halogenated hydrocarbons and polyaromatic hydrocarbons have been shown to be mutagens, carcinogens, or both. Because some of these substances are released into the atmosphere (*1*), it is possible that communities surrounding petroleum and chemical plants are placed at increased risk of cancer and other adverse health outcomes.

Previous studies have observed associations between residence in petroleum and chemical manufacturing counties and cancer mortality rates (*2–4*). Workers in the petroleum industry have been reported to be at increased risk for cancer of the stomach, liver and biliary passages, pancreas, esophagus, brain and skin, and for leukemia and multiple myeloma, although findings have not been consistent from study to study (*5–8*). Excess risk of death from cardiovascular disease has also been noted among workers in petroleum refining and petrochemical plants (*6*). However, the long-term health effects of petroleum and chemical industry air emissions on surrounding communities have not been well studied.

Since the beginning of the century, the north and west portions of Contra Costa County have been the center of the petroleum refining and chemical manufacturing industries in northern California, and there has recently been public concern over possibly elevated cancer rates in the county. In the present study, a model of air dispersion was used to partition the county by degree of exposure to emissions from petroleum and chemical plants. We then examined the degree of association between exposure and two groups of health outcomes, cancer incidence and mortality from major causes of death. Census socioeconomic status (SES) data were obtained for the county, and used to control for the potentially confounding effects of SES.

Methods

Apart from the endpoints being considered, the methods used for the cancer incidence and major-cause mortality studies were identical. The county was divided into four areas by level of residential exposure to air pollutants emitted by major industries, and SES measurements from the U.S. Census were obtained for the areas. Then cancer incidence rates and major-cause mortality rates were calculated for the four areas, and the degree of association between the rates and the level of exposure to emissions from petroleum and chemical plants was evaluated. For both cancer incidence and mortality, we analyzed rates among the white population only, because the proportion of other racial groups is small [18.4% in the 1970 Census (*9*) and 15.7% in the 1975 Census (*10*)] and unevenly distributed across exposure areas.

Estimation of Residential Exposure to Industrial Air Emissions

We wished to assess exposure levels to the air emissions from petroleum and chemical plants in Contra Costa County. In previous studies of the health effects of industrial emissions, the county was divided either according to the presence or absence of industry (*11*) or according to monitoring station levels (*12*). Because of the proximity of the petroleum and chemical plants to other industries and major highways, neither of these systems specifically defines exposure to petroleum and chemical plant emissions. We have relied instead on a model developed by the Bay Area Air Quality Management District (BAAQMD).

Since 1972 the BAAQMD has been estimating the quantities of sulfur dioxide (SO_2), hydrocarbons (HC), and oxides of nitrogen (NO_x) emitted from all major industries in the county. These emission estimates are derived from measurements made at the point of emission, combined with information on quantities of chemicals used and produced by the industries. The most complete estimates during the observation period of our study were made in 1975. The BAAQMD also estimates quantities of these and other pollutants emitted by nonindustrial sources such as automobiles, aircraft and small businesses.

For the purposes of this study, the BAAQMD developed a pollution dispersion model which used as inputs the industrial emission estimates for 1975, combined with topographic data from the United States Geologic Survey and meteorologic data obtained from measurements taken at BAAQMD monitoring stations in 1973. Emissions produced by automobiles and other nonindustrial sources were not considered. The meteorological variables consisted of the spatial distribution of average, hourly wind speeds and the annual distribution of inversion base heights. Using the Hanna-Gifford approach to air dispersion calculations (*13*), the model generated average annual ambient concentrations of the above industrial emittants for each one square kilometer area of the county. Acute air pollution episodes such as emergency release of gases were not weighted by the model.

The basic units of study were 21 grouped census tracts. These were obtained by consolidating the 107 census tracts in Contra Costa County into 21 groups, merging census tracts that were geo-

graphically contiguous and demographically similar according to the criteria of mean family income, age distribution and race. The 21 groups were created without prior knowledge of cancer incidence or mortality rates or industrial air pollution exposures. Using population density maps and the BAAQMD-modeled industrial air pollutant concentrations for each square kilometer in the county, we calculated mean annual exposure to SO_2, HC and NO_x for each of the 21 groups for the year 1975. For example, if over the year 100 people living in one square kilometer had an average exposure to SO_2 of 10 μg/m³/day and 50 people in another square kilometer were exposed to an average SO_2 level of 22 μg/m³/day, then the mean daily exposure for the two square kilometers was 14 μg/m³/day. Thus, if there was a region made up of a large, unpopulated area with high pollution and a small, densely populated area with low pollution, the calculated mean daily exposure for the region was closer to that found in the densely populated area. Table 1 presents the estimated mean exposure to industrially emitted SO_2, HC and NO_x for each of the 21 groups.

According to the 1975 BAAQMD emissions inventory, 180 tons/day of HC and 130 tons/day of SO_2 were emitted by all sources. Petroleum refineries and chemical manufacturing plants emitted 68.8 tons/day of HC and 92 tons/day of SO_2 and thereby accounted for 95% of the HC and 75% of the SO_2 emitted from major industries in Contra Costa (*14*). Power plants accounted for another 20% of the industrially emitted SO_2, but less than 1% of emitted HC. Our model therefore assumes that the presence of industrially emitted HC and SO_2 in the air of a grouped tract is a marker for exposure to gaseous emissions from the petroleum and chemical industries. Conversely, if either HC or SO_2 is absent we consider that the location is unexposed to emissions from these industries. The BAAQMD estimates are only broadly indicative of exposure levels rather that quantitatively precise. Therefore, the 21 grouped tracts were further consolidated into four areas,

Table 1. Average exposures to industrially produced sulfur dioxide, hydrocarbons and oxides of nitrogen by grouped census tracts in Contra Costa County, California, 1975.

Exposure area (see Fig. 1)	Group	Census tracts[a]	SO_2 μg/m³	Hydrocarbons, μg/m³	Oxides of nitrogen, μg/m³
4	1	315, 316, 317, 318, 319, 320, 357, 358	3.5	3.3	2.8
4	2	327, 328, 329, 330	1.8	1.4	0.5
4	3	362, 370, 377, 378, 379, 380, 381, 382, 383, 384, 386	2.5	2.9	1.8
4	4	369, 371, 372, 374, 375	1.0	2.9	1.4
4	5	365, 366, 367, 368, 373, 376	2.7	5.8	4.9
3	6	356, 359, 360	0.4	0.9	0.7
3	7	309, 310, 311, 312, 314	0.6	0.6	2.4
3	8	321, 322, 323, 324	0.5	0.8	0.7
2	9	361, 363, 364, 385, 387, 388, 389, 390, 391, 392	0.0	0.7	1.3
2	10	307	0.0	0.0	1.6
2	11	305, 306, 308	0.0	0.0	2.3
2	12	301, 302, 303, 304	0.0	0.0	0.5
2	13	313	0.0	0.0	0.9
1	14	337, 355	0.0	0.0	0.0
1	15	331, 332, 335	0.0	0.0	0.0
1	16	333, 334	0.0	0.0	0.0
1	17	336	0.0	0.0	0.0
1	18	338, 343	0.0	0.0	0.0
1	19	344, 345, 346	0.0	0.0	0.0
1	20	342, 347, 348, 349, 350, 351, 352, 353, 354	0.0	0.0	0.0
1	21	325, 326, 339, 340, 341	0.0	0.0	0.0

[a]First three digits of census tract only.

based on degree of exposure to air emissions from major industry (see Table 1).

Area 4 consists of regions with relatively high concentrations of SO_2 and HC from industrial sources (>1 μg/m³ of each); area 3 includes areas with lower concentrations of industry-produced SO_2 and HC (0–1 μg/m³ of each); area 2 has no exposure to industrially emitted SO_2 and almost no exposure to HC. Areas 2, 3 and 4 all have exposure to industrially emitted NO_x. Thus, areas 2, 3 and 4 are exposed to industrial air emissions, but, based on the virtual absence of SO_2 and HC, area 2 is considered to be unexposed to emissions from the petroleum and chemical industries.

Area 1 has no air exposure to industrially produced SO_2, HC or NO_x; therefore, for the purpose of this study it is considered to be unexposed to any sources of industrial air emissions. Figure 1 shows the four areas and the location of major industry and highways in the county.

Although adverse health outcomes have been associated with exposure to SO_2 and HC, the levels estimated in this study are far below those at which adverse long-term health effects have been observed (*15*). However, this does not preclude the use of SO_2 and HC as markers for relative exposure to petroleum and chemical industry emissions.

Socioeconomic Status

Because rates of certain cancers and other causes of death have been associated with various socioeconomic characteristics of the population, we calculated SES measures for the exposure areas. Data from the 1970 census were used to generate, for the white population of each area, mean family income, percent blue collar workers, percent high school graduates and percent residents with Spanish surnames (*9*).

As can be seen in Table 2, these measures are similar in areas 2, 3 and 4, although area 4 has the longest mean residence time and is of slightly lower socioeconomic class. Area 1, on the other hand, has much higher values for income and education and a lower percentage of blue collar workers and Spanish-surnamed residents.

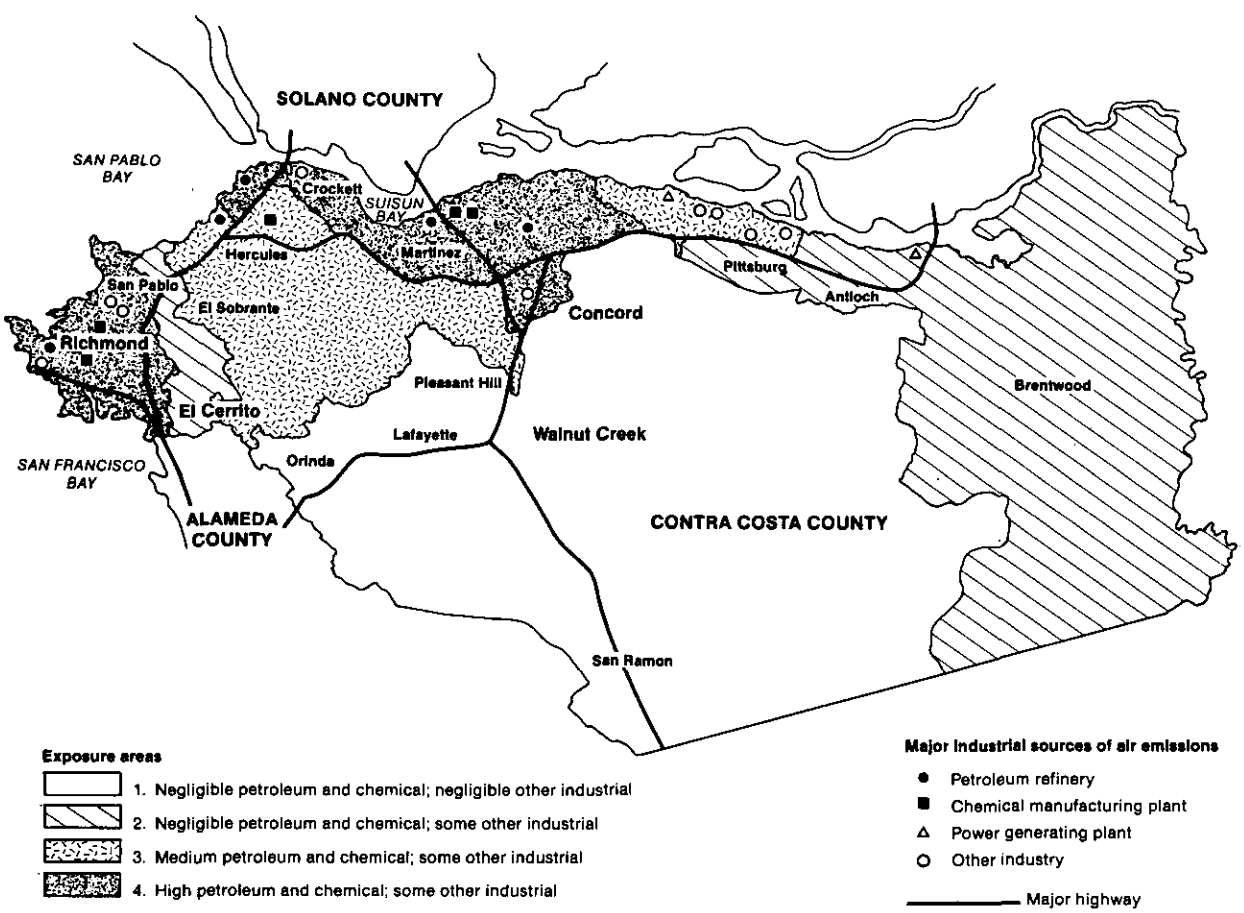

FIGURE 1. Exposure areas in Contra Costa County.

Table 2. Demographic indicators in Contra Costa County, California, by exposure area, whites only.[a]

	Area 1	Area 2	Area 3	Area 4
Mean family income ($)[b]	16,661	12,417	12,611	11,430
Blue collar, %[c]	21.0	37.9	38.1	40.3
High school graduate, %	80.9	62.4	63.4	56.4
Spanish surnamed, %	5.3	14.3	12.0	10.9
Estimated white male population at risk, 1969–1977[d]	1,032,571	411,180	300,494	486,691
Estimated white female population at risk, 1969–1977[d]	1,061,451	421,995	303,397	506,191
Mean years residente	4.70	5.75	5.10	6.33

[a]Data from 1970 Census, unless otherwise specified. The census provides information on the above measures by census tract for the population as a whole, and for the black population in census tracts with more than 400 black residents listed. Therefore, if an area contains a census tract with fewer than 400 black residents, their contribution will be included in the above data.

[b]Data are available by census tract for total mean and median income and for black median income. To obtain an estimate of white mean family income, the contribution of blacks had to be removed. For this purpose, we estimated the black mean income by adjusting the black median by the ratio between the mean and median family income for the total population of the tract.

[c]The number of craftsmen, operatives, transportation workers and laborers as a percentage of total employed.

[d]Estimated by Resource for Cancer Epidemiology, Surveillance Epidemiology End Results Program, Emeryville, California.

[e]Median years only are available for each census tract. The mean obtained for each area is actually a population-weighted average of the census tract medians, rather than a true mean. Source: 1975 Contra Costa County Special Census.

Cancer Incidence Rates

Seventeen cancer sites were selected for study on the basis of reported association in the human and animal literature with exposure to petrochemicals and other chemicals such as heavy metals and asbestos. Three additional sites which have not previously been associated with industrial chemical exposures but which have well-documented socioeconomic associations, namely, female breast, uterine corpus and uterine cervix, were also included in the study. Cancer incidence rates in the four areas of Contra Costa County for the period 1969–1977 were obtained from the Resource for Cancer Epidemiology (RCE). The RCE is part of the National Cancer Institute's Surveillance Epidemiology and End Results Program. It reviews death certificates and surveys hospitals in the five San Francisco Bay Area counties (San Francisco, Alameda, Contra Costa, San Mateo and Marin), and surveys referral hospitals in adjacent counties in order to maintain a registry of all newly diagnosed cancer cases in the Bay Area. Overall, the RCE ascertains an estimated 98% of all cancer cases, and the diagnosis for 90–95% of these cases is confirmed by pathologic report (Dr. Donald Austin, personal communication). Variables recorded include cancer site and histology, age, race, sex, date of diagnosis and census tract of residence at time of diagnosis. To provide denominator figures for rate calculations, the RCE maintains current population estimates by age, race and sex. These are derived from census data (which include the 1975 special census taken in Contra Costa County) and Department of Finance estimates for intercensal years.

Average annual age- and sex-specific incidence rates for each of the four exposure areas were calculated for the twenty cancer sites (see Table 3). Age-adjusted rates were computed by the direct method, the 1950 population of the Continental United States being used as the standard. The all-site category contains all malignancies including some which do not appear in Table 3. The sites individually specified in Table 3 constitute 86% of the all-site category.

Standard errors for age-specific and age-adjusted rates were obtained in the usual way (*16*).

Major-Cause Mortality Rates

We calculated mortality rates for the four areas as follows. A computer tape containing summary information for every death occurring to a Contra Costa County resident within the years 1968–1972 was obtained from the Vital Statistics Section, California State Department of Health Services. The tape records contained data abstracted from the death certificates including year of death, age at death, sex, race, census tract of residence, state file number and ICD cause of death for each deceased individual. In the 5-yr period under study, 17,427 deaths occurred in the white population. The individuals who had died while outside the county (approximately 3,000) were not coded for census tract of residence on the tape. We therefore made a 20% sample of these untracted deaths on the basis of the last digit of the state file number and examined the death certificates of the sampled individuals. The certificates contain the residential address, and the census tract of residence could thus be ascertained. The causes of death were grouped into eight major categories. These are: cancer, cardiovascular disease, cerebrovascular disease, re-

Table 3. Average annual age-adjusted cancer incidence rates per 100,000 by exposure area and sex, Contra Costa County, California, whites only, 1969–1977.[a]

Sex	Cancer (ICD 8 Code)	Area 1	Area 2	Area 3	Area 4	Total[b]	*p* values[c]
Male	Buccal, pharyngeal excluding nasopharynx (140.0–146.9, 148.0–149)	9.3 (87)[d]	13.2 (53)	18.9 (47)	22.4 (119)	12.6 (324)	0.0008
	Nasopharynx, nose, sinus (147, 160.0–160.9)	1.3 (12)	2.3 (9)	1.8 (5)	2.0 (11)	1.7 (37)	—
	Esophagus (150)	4.7 (39)	4.0 (15)	3.7 (9)	5.1 (27)	4.5 (93)	—
	Stomach (151)	8.4 (72)	7.4 (28)	11.0 (26)	14.0 (76)	8.4 (216)	0.002
	Colon, rectum, anus (153.0–156.9)	49.6 (420)	43.3 (167)	37.5 (89)	49.4 (268)	46.5 (989)	
	Liver (155)	2.4 (21)	2.7 (10)	3.1 (7)	4.2 (23)	3.1 (63)	—
	Pancreas (157)	8.6 (75)	9.6 (36)	8.7 (22)	10.5 (54)	9.2 (207)	—
	Trachea, bronchus, lung (162)	56.3 (492)	75.9 (298)	82.3 (195)	89.1 (479)	69.2 (1555)	0.02
	Leukemia (204–207.9)	12.1 (109)	7.6 (29)	15.2 (37)	9.2 (48)	10.3 (230)	0.09
	Melanoma (172)	12.3 (120)	8.5 (34)	9.1 (25)	5.3 (28)	8.5 (215)	—
	Prostate (185)	59.6 (474)	46.3 (171)	48.1 (103)	62.2 (334)	58.7 (1139)	0.002
	Testicle (186)	4.0 (41)	4.0 (16)	3.7 (11)	4.1 (21)	4.3 (94)	
	Bladder (188)	25.0 (213)	21.7 (81)	18.2 (42)	19.8 (105)	21.9 (464)	
	Kidney, urinary organs (189)	10.1 (89)	7.3 (28)	12.7 (32)	10.1 (53)	9.6 (217)	0.05
	Brain, nervous system (191–192.9)	5.3 (52)	8.3 (33)	9.0 (24)	8.3 (43)	7.1 (152)	—
	Hodgkins lymphoma (201)	5.3 (52)	3.1 (13)	2.9 (8)	4.5 (22)	4.1 (98)	—
	Other lymphoma (202, 202.0–202.9)	9.6 (90)	8.5 (32)	7.2 (18)	9.4 (50)	9.0 (199)	—
	All sites (140.0–239.9)	325.7 (2827)	309.7 (1194)	331.4 (793)	371.0 (1978)	348.9 (7162)	<0.0001
Female	Buccal, pharyngeal excluding nasopharynx (140.0–146.9, 148.0–149)	6.7 (71)	5.2 (25)	8.5 (24)	7.9 (50)	6.8 (176)	0.04
	Nasopharynx, nose, sinus (147, 160.0–160.9)	0.9 (9)	0.5 (2)	0.4 (1)	0.8 (5)	0.7 (18)	—
	Esophagus (150)	1.6 (16)	2.4 (11)	0.4 (1)	1.9 (13)	1.3 (43)	0.05[e]
	Stomach (151)	4.8 (57)	7.1 (31)	6.3 (18)	6.2 (42)	5.6 (155)	—
	Colon, rectum, anus (153.0–154.2)	36.8 (413)	38.3 (168)	39.4 (109)	39.0 (261)	38.0 (982)	—
	Liver (155)	1.6 (11)	0.9 (4)	0.4 (1)	1.9 (12)	1.2 (29)	—
	Pancreas (157)	6.7 (71)	7.2 (33)	6.0 (16)	7.7 (56)	7.3 (185)	—
	Trachea, bronchus, lung (162)	21.2 (223)	22.9 (103)	28.5 (79)	21.5 (186)	22.3 (618)	—
	Leukemia (204–207.9)	5.5 (58)	7.0 (29)	6.9 (19)	6.9 (42)	6.2 (151)	—
	Melanoma (172)	9.7 (105)	8.0 (33)	12.4 (35)	6.8 (38)	8.7 (226)	—
	Breast (185)	93.3 (1025)	81.2 (355)	83.2 (241)	83.2 (506)	87.0 (2219)	—
	Cervix uteri (186)	8.9 (97)	15.2 (63)	14.7 (41)	18.0 (98)	11.6 (316)	—
	Corpus uteri (187)	49.7 (504)	36.1 (160)	34.8 (98)	35.3 (225)	40.6 (1030)	—

Table 3 (Continued)

Sex	Cancer (ICD 8 Code)	Area 1	Area 2	Area 3	Area 4	Total[b]	p values[c]
Female	Bladder (188)	6.7 (74)	6.6 (30)	2.2 (6)	5.7 (39)	4.7 (156)	0.05[e]
	Kidney, urinary organs (189)	4.3 (46)	4.2 (18)	2.2 (6)	5.2 (30)	4.0 (106)	
	Brain, nervous system (191–192.9)	4.1 (45)	3.7 (16)	6.1 (17)	4.9 (27)	5.7 (105)	—
	Hodgkins lymphoma (201)	2.4 (27)	2.8 (13)	2.2 (8)	3.2 (17)	0.4 (65)	—
	Other lymphoma (202, 202.0–202.9)	5.5 (57)	5.1 (22)	4.1 (11)	5.3 (33)	5.7 (130)	—
	All sites (140.0–239.9)	316.8 (3432)	298.5 (1309)	301.3 (852)	317.0 (1972)	320.1 (7891)	0.07

[a]Age-adjusted rates computed by direct method, using the 1950 total population of the Continental United States as the standard.
[b]Includes cases not assigned to a particular area.
[c]p values are reported at $p < 0.1$ for positive trend test among areas 2, 3 and 4.
[d]Number in parentheses is total number of cancer cases, 1969–1977.
[e]Negative trend significant at $p < 0.1$

spiratory disease, cirrhosis, metabolic disease, violence and other. The ICD 8 codes included in each grouping are indicated in Table 1. The first five categories were chosen because they have been related in the animal and human literature to exposure to specific chemical agents, some of which are emitted by petroleum and chemical plants. The categories of metabolic disease (primarily diabetes) and violent deaths (which include accidents, suicides and poisonings) serve as "controls", in that we would expect no relationship between them and industrial emissions.

We then computed average annual age-adjusted and age-specific mortality rates for the eight categories by sex and exposure area. Numerators for age-specific rates were obtained by adding the number of deaths originally tracted to the number estimated to be untracted by the 20% sample of untracted deaths in the appropriate age by sex by cause by area category.

The 1970 United States Census provided population denominators. Standard errors for the rates were also estimated, taking into account the 20% sampling (see Appendix 1). Rates were adjusted by the direct method to the overall white age distribution in the county.

Statistical Evaluation of Trends in Rates

Areas 2, 3 and 4 are comparable with respect to SES, while area 1 is of much higher SES. We therefore assessed the statistical significance of the gradients in cancer incidence and mortality rates with level of exposure to petroleum and chemical industry emissions using only the areas 2, 3 and 4.

The $\bar{\chi}^2$ test (*17*) was used to test for increasing (and decreasing) trends in rates. The test assumes that rates are normally distributed and tests the null hypothesis of equal rates against the alternative of increasing (or decreasing) rates between at least areas 2 and 3 or 3 and 4. Although area 1 rates were not used in the test because of the possibility of confounding, they were examined to check for consistency with any pattern observed among areas 2, 3 and 4. Significance levels are only reported if areas 3 and 4 both have higher rates than area 2 and if the p value for increasing trend is less than 0.1 or if area 2 has a higher rate than both areas 3 and 4 and the p value for decreasing trend is less than 0.1.

Results

Cancer Incidence

Table 3 reports average annual age-adjusted cancer incidence rates for areas 1, 2, 3 and 4 by sex for the years 1969–1977. The p values for the statistical test for trend among areas 2, 3 and 4 are also provided. For males there was a statistically significant ($p < 0.05$) increasing trend in cancer incidence from area 2 to area 4 for the following cancers: buccal cavity and pharyngeal excluding nasopharynx; stomach; combined trachea, bronchus and lung; prostate; combined kidney and urinary organs; and all-site. For females there was a statistically significant increasing trend ($p < 0.05$) only for buccal cavity and pharyngeal cancer; there were significantly decreas-

ing trends for cancer of the esophagus and the bladder.

As expected, cancers of the female breast and uterine corpus have similar rates in areas 2, 3 and 4 and higher rates in area 1, and rates of cancer of the uterine cervix are lower in area 1 compared to areas 2, 3 and 4. Similarly, among males, area 1 has generally lower incidence rates for those cancers with an inverse SES gradient (lung, buccal cavity and pharyngeal, esophageal, stomach and liver); and higher rates for those cancers positively associated with SES (Hodgkins disease, melanoma, leukemia, colon, rectal, prostatic, and testicular) (*18*).

For those cancers where the positive trend was significant, trends in age-specific rates were statistically significant for a number of age groups among persons 40 years and older (results not shown).

Major-Cause Mortality

Table 4 reports the age-adjusted mortality rates for each of the four exposure areas by sex. It can be seen that there are significant ($p < 0.05$) increasing trends among the three areas for deaths from cardiovascular disease and total causes in both sexes, and from cancer and cirrhosis in men. The trend for cancer and cerebrovascular disease mortality rates in women was also declared significant at the 0.05 level by the trend test, although rates for both these causes were slightly lower in area 3 than in the other two areas.

Age-adjusted death rates in area 1 are equal to or lower than those in areas 2, 3 and 4 for all categories, with the exception of cerebrovascular disease in both sexes, and cancer in women. For the categories where there was an increasing trend among areas 2, 3 and 4, the trends in the age-adjusted rates are generally present in several age groups over 40, although statistical significance was not achieved in all cases (results not shown).

Discussion

Our results show a positive relationship between estimated residential exposure to petroleum refinery and chemical plant air emissions and incidence rates for several cancers for the years 1969–1977. The effects are most prominent in men, though for cancers of the buccal cavity and pharynx there is a gradient of risk in both males and females. We have also found an association between exposure and mortality from a number of major causes of death for the years 1968–1972 in both men and women.

For cancer sites whose p value for increasing trend was less than 0.1, Table 5 compares average annual incidence rates in areas 3 and 4 (the areas exposed to petroleum and chemical emissions) with age-adjusted rates for the San Francisco–

Table 4. Average annual age-adjusted mortality rates per 100,000 by exposure area and sex, Contra Costa County, California, whites only, 1968–1972.[a]

Sex	Cause of death (ICD 8 Code)	Area 1	Area 2	Area 3	Area 4	Total	p values[b]
Male	Cancer (140–239)	125.9	151.8	159.8	179.1	149.6	0.034
	Cardiovascular (390–403, 440–458)	273.0	321.7	373.5	395.6	328.3	0.0002
	Cerebrovascular (430–438)	57.3	55.8	53.0	67.3	59.8	0.055
	Respiratory (460–519)	44.4	60.2	62.9	55.6	53.2	—
	Cirrhosis (571)	11.4	20.7	22.8	31.1	20.0	0.012
	Violent (810–999)	79.7	128.1	116.2	112.7	101.0	—
	Metabolic (240–279)	10.4	10.3	10.7	10.2	11.6	—
	Other	48.2	53.8	72.9	99.7	66.0	<0.0001
	All	650.4	802.4	871.8	951.4	789.4	<0.0001
Female	Cancer (140–239)	123.1	123.3	107.2	147.6	129.3	0.003
	Cardiovascular (390–403, 440–458)	203.7	212.6	240.5	263.5	228.2	0.0006
	Cerebrovascular (430–438)	85.5	66.1	60.3	101.7	84.9	<0.0001
	Respiratory (460–519)	26.2	26.0	30.2	31.6	28.9	—
	Cirrhosis (571)	10.2	12.5	16.6	16.1	12.9	—
	Violent (810–999)	43.8	45.8	49.2	50.8	47.4	—
	Metabolic (240–279)	7.8	13.9	15.4	16.6	13.3	
	Other	40.7	61.2	67.0	71.4	55.9	—
	All	541.0	561.4	586.5	699.2	600.7	<0.0001

[a]Age-adjusted rates computed by the direct method, using the total population of Contra Costa County at the 1970 U.S. Census as the standard.

[b]p values are reported if $p < 0.1$ for increasing trend among areas 2, 3 and 4. No negative trends were significant at $p < 0.1$.

Table 5. Average annual age-adjusted cancer incidence rates per 100,000 in areas 3 and 4, Contra Costa County, California, 1969–1977, compared to San Francisco–Oakland SMSA and United States cancer incidence rates.[a,b]

Sex		Area 3	Area 4	SF–Oakland SMSA 1969–1971	United States[c] 1969–1971	SF–Oakland SMSA[d] 1973–1977
Male	Buccal, pharyngeal excluding nasopharynx	18.9	22.4	17.3	15.1	17.2
	Stomach	11.0	14.0	14.2	12.0	12.4
	Trachea, bronchus, lung	82.3	89.1	68.4	64.3	74.5
	Leukemia	15.7	9.2	10.6	11.4	10.5
	Prostate	48.1	62.2	49.3	45.9	54.7
	Kidney, urinary organs	12.7	10.1	8.8	8.6	9.5
	All sites excluding *in situ*	331.4	371.0	330.1	300.9	344.8
Female	Buccal, pharyngeal exluding nasopharynx	8.5	7.9	7.5	4.9	7.9
	All sites	301.3	317.0	291.0	252.0	325.0

[a]Age-adjusted rates computed by the direct method, using the 1950 total population of the Continental United States as the Standard.

[b]San Francisco–Oakland includes Contra Costa County, San Francisco County, Alameda County, Marin County and San Mateo County.

[c]Combined Standard Metropolitan Statistical Areas used in the Third National Cancer Survey.

[d]Surveillance Epidemiology End Results Program.

Oakland Standard Metropolitan Statistical Area (SMSA) for 1969–1971 and 1973–1977 and with age-adjusted rates for the combined SMSAs of the Third National Cancer Survey (*19,20*). For cancer of the buccal cavity and pharynx, stomach, combined trachea, bronchus and lung, combined kidney and urinary organs and all-site, incidence rates for males residing in area 4 were higher than rates in the San Francisco–Oakland SMSA and in the combined SMSAs. Differences between area 3 and the other SMSAs are not so striking, and the male stomach cancer rate in that area is actually lower than the comparison stomach cancer rates.

The cancer incidence results are consistent with the findings of a number of previous studies. Blot and Fraumeni (*2*) examined age-adjusted mortality rates for lung cancer in all United States counties from 1950 to 1969 and found high rates in men in counties where paper, chemical, petroleum and transportation industries were located. Blot et al. (*3*) found excess mortality from cancer of the lung, nasal cavity, sinuses and skin among male residents of United States counties where the petroleum industry is most heavily concentrated. Hoover and Fraumeni (*4*) compared average annual age-adjusted cancer mortality rates from 1950 to 1969 between whites residing in chemical industry counties and whites residing in other United States counties. They found excess mortality for both sexes among whites residing in chemical industry counties for total cancers, bladder cancer, liver and gallbladder neoplasms, cancer of the nasopharynx and nasal sinuses and malignant melanoma. These results were not readily explained by confounding due to degree of urbanization, socioeconomic class or employment in non-chemical industries. Finally, Gottlieb and co-workers (*21*) used a case-control design to study lung cancer mortality in Louisiana and found an association with residential proximity to petroleum and chemical industries.

The results of our study are in contrast to the prior work of Hearey et al. (*11*), who found no relationship between cancer incidence and residential exposure to petroleum and chemical emissions in Contra Costa County. These authors, however, defined their exposure area in Northern Contra Costa County in a somewhat arbitrary manner, in that a roughly east-west line was drawn to separate the industrial sector from the remainder of the county. If exposure is defined more accurately in our study, then the absence of differences between rates in exposed and control groups in Hearey et al. (*11*) might be a result of misclassification (*22*).

Another possibly important difference between the two studies lies in the source of incident cases. Hearey et al. (*11*) based their rates on a 10% sample of Kaiser Permanente Medical Plan members, while our study employed a nearly complete case ascertainment system. Because of these methodologic differences, it would be of interest to determine whether the associations we have observed can be reproduced within the Kaiser population if the exposure areas used in the present study are employed.

In an as yet unpublished case-control study, Austin et al. (*12*) found that lung cancer risk in Contra Costa County was related to occupation and related to residential exposure to industrial emittants only through the confounding effect of

occupation. Similarly, Henderson et al. (*23*) found an excess incidence of lung cancer among males in south central Los Angeles, which was initially hypothesized to be due to an increased level of atmospheric polyaromatic hydrocarbons. However, when the same group followed up their findings with a case-control study, they concluded that the previously noted differences in lung cancer rates were probably explained by differences in occupational and smoking patterns (*24*).

It is difficult to check the consistency of the major-cause mortality findings with other studies, as the effect of petroleum and chemical industry emissions on mortality in nearby residential areas does not appear to have been previously studied. A number of studies have examined the relationship between measured air quality variables and mortality, and the results have suggested that elevated death rates from various cancers and cardiovascular disease coincide with higher levels of ambient sulfur dioxide [see Lave and Seskin (*15*) for a comprehensive review of the literature up to 1977].

Inferences of causality from observational studies, in particular from ecologic studies of the kind described here, are potentially subject to a number of sources of bias. These include multiple significance tests, inaccurate definition of exposure variables, the "ecologic fallacy," and confounding due to unmeasured variables. We discuss the extent to which each of these sources may be operative in this study.

Whenever multiple significance tests are conducted, some will be expected to have statistically significant results by chance alone. In our analysis, if there was no effect on the risk of cancer at any site produced by petroleum and chemical plant air emissions, we would expect that: (1) only two or three cancers out of 20 would have increasing trends with $p < 0.10$; (2) all-site cancer incidence would be similar among areas 2, 3 and 4; and (3) there would be an equal number of statistically significant increasing and decreasing trends across areas 2, 3 and 4. For males, none of these three possibilities is realied. Six cancers show an increasing trend at $p < 0.10$, all-site cancer incidence increases significantly as estimated exposure to petroleum and chemical air emissions increases, and there is no cancer site for which the age-adjusted rates have a decreasing trend with $p < 0.10$. For female cancer incidence rates the evidence against the multiple significance test explanation is less convincing. While there is a weak increasing trend in all-site cancer incidence rates ($p = 0.07$), there is only one cancer site with a significantly increasing trend at $p < 0.05$ (buccal cavity and pharynx), and there are two cancers whose rates are greater in area 2 than in either area 3 or area 4, and for which there are significant decreasing trends. The positive trend for buccal cavity and pharynx cancer could therefore have occurred by chance alone, but this explanation is less likely since there is also a positive trend in this cancer for males.

Of the 16 independent tests for increasing trend in age-adjusted mortality rates, 7 were significant at less than the 0.05 level, as compared with the one we would expect on the basis of random fluctuations if there were no true trends among the three areas for any cause of death. Moreover, when we tested for decreasing trend none of the gradients was significant.

The estimates of residential exposure levels in this study are generated by means of a model, which in effect defines exposure by proximity to SO_2 and HC-emitting industry, with some weighting applied to allow for wind direction and topography. This model was used in preference to monitoring station data, to enable us to separate the petroleum and chemical plant emissions of interest from other major industrial, small business and automobile air emissions. The levels of exposure estimated by the model cannot be precisely verified. However, they are consistent with the expectation that locations closer to and downwind from petroleum and chemical plants should be classified as having higher exposures than locations farther away from the plants (see Fig. 1). Furthermore, the pattern of estimated exposure levels is similar to that of monitored levels of SO_4 in Contra Costa County (*25*). A potential problem with the model is that 1975 emissions were used, whereas the relevant exposures may have occurred 20–30 years earlier. We have assumed that the qualitative relationship between exposure levels in areas 2, 3 and 4 has been relatively constant in the recent past.

In this study, the mean exposure and SES levels of an area are assumed to apply to all individuals within the area. This assumption could obscure differences in confounding variables which might be responsible for the observed differences in rates, giving rise to the so called "ecologic fallacy". For example, an explanation of our observations could run as follows: a high and a low SES subgroup in area 4 produced average SES levels for the whole area equal to another area, say area 3, which had only mid-level SES residents. Then, if there was a nonlinear relationship between SES and cancer incidence or mortality rates such that high and middle SES groups were

at equal risk, but the risk of low SES groups was elevated, the differential rates between area 3 and area 4 would result. This difference in cancer incidence or mortality rates would be ascribed to the difference in exposure to petroleum and chemical industry emissions, when in fact it was due to SES differences. However, the presence of gradients across more than two exposure levels decreases the likelihood that the "ecologic fallacy" can explain our observations. On the other hand, it certainly does not rule it out.

A number of potentially confounding variables were not measured in this study. The most important of these are occupation, smoking status, level of alcohol consumption and other industrial air pollution exposures.

Occupation can be an important confounder in any study of residential exposure, because people tend to live near their worksite. If area of residence alone produced the effect on cancer incidence rates observed in our study, we would have seen similar gradients in the incidence rates of both sexes. Our results, however, are much more dramatic in men. Moreover, most of the cancer sites for which significant increasing trends were observed in men have been associated with occupational exposures in the petroleum and chemical industries, and in other industries such as ship building, which have been present in Contra Costa County. Cancer of the buccal cavity and pharynx (essentially, of the mouth and cheek), for which there was a relationship in both sexes, has not been previously associated with occupational exposures in these industries. Although areas 2, 3 and 4 used in the statistical analysis were reasonably homogeneous for several socioeconomic indicators, including percentage of blue collar workers, it is still possible that the portion of blue collar workers occupationally exposed to carcinogens might have varied among the areas. For major-cause mortality, the fact that we observed similar trends in men and women in deaths from cardiovascular and cerebrovascular disease argues against a uniquely occupational explanation of these outcomes.

It might also be argued that the observed gradients in rates are due to differences in lifestyle variables such as tobacco and alcohol consumption among areas 2, 3 and 4. Either one or the other of these two factors have been associated with most of the cancers for which there is an increasing trend among areas 2, 3 and 4 (*18*). Similarly, cardiovascular and cerebrovascular disease mortality has been associated with tobacco consumption, and alcohol consumption is one of the strongest known risk factors for cirrhosis. In order for these variables to have a confounding effect, their distribution must coincide with the distribution of estimated petroleum and chemical plant emissions. There is unfortunately no information available on smoking and alcohol consumption patterns within the subareas of Contra Costa County.

Another potentially confounding variable is exposure to emissions from industry other than petroleum and chemical plants. Many such industries are or were located in areas 3 and 4, and it is possible that emissions from these sources are the cause of the trends observed in this study.

Although we do not have data on important confounding variables, we can use a model similar to that discussed by Schlesselman (*26*) to estimate the relative risk that unmeasured confounding variables must confer in order to account for the observed differences between the rates of area 2 and area 4 (see Appendix 2). These estimated relative risks are displayed in Table 6 for those cancer sites and causes of death where the p value for increasing trend was less than 0.1 and under the assumption that the confounding risk factors (all potential factors, including occupation, smoking and alcohol consumption and other past or present industrial emissions) have a prevalence of 50% in area 4 and 35% in area 2. Under these assumptions, covariate bias could not totally explain the difference in cancer incidence rates between area 2 and area 4 for cancer of the buccal cavity and pharynx, stomach and kidney and urinary organs in men, or for buccal cavity and pharyngeal cancer in women. The other differences in cancer incidence rates be-

Table 6. Estimated relative risks for a dichotomous variable with prevalence 50% in area 4 and 35% in area 2 if it alone is responsible for the observed difference in rates between the two areas: selected categories.

	Male	Female
Cancer incidence		
Buccal, pharyngeal	a	a
Stomach	a	
Prostate	12.52	
Trachea, bronchus, lung	2.95	
Kidney, urinary organs	25.23	
All sites	3.45	1.48
Mortality		
Cancer	3.1	3.4
Cardiovascular	4.3	4.6
Cerebrovascular	3.6	a
Cirrhosis	a	a

[a]Indicates that, according to our estimation procedure, the unmeasured dichotomous variable with prevalences as above could not alone explain the observed differences between the rates in areas 4 and 2.

tween areas 2 and 4 could be explained by covariable(s) conveying relative risks between 2.95 and 12.52. For mortality, the effect on cerebrovascular disesase mortality in men could not be explained by a covariate with the postulated prevalences. The differentials in other causes of death could arise with covariates which conferred a relative risk of between 3.1 and 4.6.

A 20–30 year latency period is usually postulated between initial exposure to a carcinogen and cancer diagnosis, and the causes of death of principal interest here also have a long period of initiation and development before leading to death. However, in 1975 the average lengths of residence in areas 2, 3 and 4 were 5.75, 5.10 and 6.33 yr, respectively. Therefore, in order for the petroleum and chemical air emissions to have produced the observed effects, we must postulate that either subpopulations existed within areas 3 and 4 which had long-term exposures to petroleum and chemical air emissions, or the effect was one of "promotion" rather than "initiation." The slightly longer mean residence time for area 4 as compared with residence times in areas 2 and 3 is evidence in favor of the first alternative.

In conclusion, we have observed associations between cancer incidence and major-cause mortality rates and estimated residential exposure to petroleum and chemical industry air emissions. Because the cancer incidence associations were far stronger in men, they may be attributable to occupational factors or smoking and alcohol usage. Similarly, while the mortality results cannot be easily attributed to occupation alone, they may be due to unmeasured lifestyle variables, such as tobacco and alcohol consumption, or socioeconomic differences among the areas. We conclude that further studies based on individuals should be carried out in which more direct control of potentially confounding factors will be possible. Until such studies are carried out, the implications of our findings for the population of Contra Costa County, and for other populations located near petroleum refineries and chemical plants, must remain uncertain.

Appendix 1

The variances of mortality rates are calculated in the usual way for age-specific and age-adjusted rates (*16*), except that when numerators were obtained partly from a 20% sample of untracted deaths as described in the text, the additional variability thereby incurred was incorporated in the estimate of variance of the rate in each cause by age by sex by area cell. This was done as follows: Suppose for a given such cell, with a population N, there are T deaths tracted, U untracted and a 20% sample of untracted deaths for that cause by age by sex group yields U_S deaths in the cell.

Then our estimate of the cell specific death rate $\hat{p}$ would be

$$\hat{p} = (T + 5U_S)/N$$

This has variance $p_T(1 - p_T)/N + 25V(U_S)/N^2$, where p_T is the probability of dying in the cell and being tracted. $V(U_S)$ is obtained via the iterated formula for variance,

$$V(U_S) = V[E(U_S|U)] + E[V(U_S|U]$$

Here V, E and "|" denote variance, expectation and conditional on, respectively. Then, if p_U is the probability of dying in the cell and being untracted and the sample used to obtain U_S may be assumed to be simple random, we may apply the simple random sampling formulae for mean and variance to give $E(U_S|U)$ and $V(U_S|U)$, as functions of $V(U_S)$ under the assumption that U is binomial (N,p_U). For the required estimate of $V(\hat{p})$, we use the estimates $\hat{p}_T = T/N$ and $\hat{p} = 5U_S/N$ in place of p_T and p_U.

Appendix 2

To estimate the relative risk of an unmeasured confounder, under the assumption that it alone accounts for the observed differences between the rates of area 4 and area 2, we assumed that a fraction f of the people in area 4 and a fraction g of the people in area 2 are exposed to some risk factor for the outcome under consideration, where $f > g$. We may think of this factor as a single agent, or as a combination of agents, but for simplicity of analysis, we assume it is dichotomous. The risk factor is then a potential confounding variable in the sudy, since it is associated with both the variable defining the exposure level to petrochemical emissions (namely, area of residence) and the cancer incidence or mortality rate (since it is a risk factor). Then, if area of residence has no effect on the risk of outcome [or $R_A = 1$ in the notation of Schlesselman (*26*)], we obtain

$$E(O_4) = Rpf + p(1 - f)$$
$$E(O_2) = Rpg + p(1 - g)$$

where $E(\ )$ denotes expected value, O_4 is the observed rate in area 4, O_2 is the observed rate in area 2, p is the probability of a person not exposed to the factor being observed as a case, and R = the relative risk of being exposed to the factor.

We then estimate R as

$$\hat{R} = [(1 - g)\,Q - (1 - f)\,]\,/\,(f - gQ) \qquad (1)$$

where $Q = O_4/O_2$.

This calculation is still appropriate if the observed rates are age-adjusted, as long as they are adjusted to the same standard, the relative risks are equal for all age groups used in the adjustment, and the fraction of each group exposed is the same within an area. The third assumption is clearly the hardest to satisfy, if we are thinking of such confounding variables as occupational exposure or cigarette smoking. However, it may be reasonable to assume some "average" fraction for the age groups which contribute to the age-adjusted rate.

A consequence of equation (1) is that $\hat{R}$ will be negative if $f/g < Q$; i.e., if the proportion exposed to the confounding variable in area 4 is less than Q times the proportion in area 2. Our interpretation of a negative estimate is that a confounder with the prevalences f and g in areas 4 and 2, respectively, could not alone account for the observed difference in rates [see discussion in Schlesselman (*26*) following his equation (*2*)].

It is important to emphasize that the relative risk we have estimated is that required to totally account for the observed difference in rates between the two areas. Because this difference is in fact an estimate of some underlying "true" difference in rates, it has an associated standard error, which can be used to construct a confidence interval for the true difference. A more conservative (lower) estimate of R would be obtained by using the lower end of this confidence interval in place of Q in equation (1).

It should also be noted that if we use $f + h$ and $g + h$ instead of f and g, where h is a positive constant, a larger estimate of R results, and that the estimate of R is inversely proportional to $f - g$. Thus, the estimate of R is conservative compared to the R which would be estimated with larger prevalences f and g, or a smaller difference between the prevalences.

This study could not have been carried out without the generous cooperation of Dr. Donald F. Austin, Resource for Cancer Epidemiology Section, California State Department of Health Services, who provided cancer incidence data; Dr. William Clark, formerly Chief, Epidemiological Studies Section, California State Department of Health Services; Dr. Edward Alpen, Director, Donner Laboratory, University of California, Berkeley; and Dr. John Goldsmith, Professor of Epidemiology, Ben Gurion University of Negev, Israel.

Financial support was provided by Grant number W-7405-ENG-48 from the Department of Energy, under Subcontract number 305-7102 from the University of California, Lawrence Berkeley Laboratory, and in part by Contract Number N01-CP-81018 from the National Cancer Institute.

REFERENCES

1. Enviro Control Inc. Industrial Hygiene Characterization of Petroleum Refineries. National Institute for Occupational Safety and Health, Cincinnati, 1979.
2. Blot, W. J., and Fraumeni, J. F. Geographic patterns of lung cancer; industrial correlations. Am. J. Epidemiol. 103: 539–550 (1976).
3. Blot, W. J., Brinton, L. A., Fraumeni, J. F., and Stone, B. J. Cancer mortality in U.S. countries with petroleum industries. Science 198: 51–53 (1977).
4. Hoover, R., and Fraumeni, J. F. Cancer mortality in U.S. countries with chemical industries. Environ. Res. 9: 196–207 (1975).
5. Hanis, N. M., Stavraky, K. M., and Fowler, J. L. Cancer mortality in oil refinery workers. J. Occup. Med. 21: 167–174 (1979).
6. Thomas, T. L., Decoufle, P., and Moure-Erasa, R. Mortality among workers employed in petroleum refining and petrochemical plants. J. Occup. Med. 22: 97–103 (1980).
7. Theriault, G., and Goulet, L. A mortality study of oil refinery workers. J. Occup. Med. 22: 367–370 (1979).
8. Gottlieb, M. S. Lung cancer and the petroleum industry in Louisiana. J. Occup. Med. 22: 384–388 (1980).
9. U.S. Bureau of the Census. U.S. Census of Population, 1970, Characteristics of the Population. Government Printing Office, Washington, DC, 1973.
10. Contra Costa County Planning Department. Census of the Population, 1975, Social Indicator Framework. Martinez, Contra Costa County, 1978.
11. Hearey, C. D., Ury, H., Siegelaub, A., Ho, M. K. P., Salomon, H., and Cella, R. L. Lack of association between cancer incidence and residence near petrochemical industry in San Francisco Bay Area. J. Natl. Cancer Inst. 64: 1295–1299 (1980).
12. California Department of Health Services. Resource for Cancer Epidemiology Section. Epidemiological study of the incidence of cancer as related to industrial emissions in Contra Costa, California, Final Report to the Environmental Protection Agency (1982).
13. Robinson, L. H. Computation of pollutant concentrations on an average annual basis. SIMS Technical Report Number 2, Bay Area Air Quality Management District, San Francisco, in press.
14. Bay Area Air Quality Management District. Base Year 1975 Emissions Inventory (1976).
15. Lave, L. B., and Seskin, E. P. Air Pollution and Human Health. Johns Hopkins University Press, Baltimore, 1977.
16. Chiang, C. L. U.S. Department of Health, Education and Welfare Vital Statistics-Selected Reports. Vol. 47, No. 9. U.S. Government Printing Office, Washington, DC, 1961.
17. Barlow, R. E., Bartholomew, D. J. Bremner, J. M., and Brunk, H. D. Statistical Inference Under Order Restrictions. John Wiley and Sons, New York, 1972.
18. Schottenfeld, D., and Fraumeni, J. F. Cancer Epidemiology and Prevention. W. B. Saunders, New York, 1982.
19. National Cancer Institute. Third National Cancer Survey: Incidence Data, Monograph No. 41. National Institutes of Health, Bethesda, 1975.
20. National Cancer Institute. Surveillance Epidemiology End Results: Incidence and Mortality Data, 1973–1977, Monograph No. 57. National Institutes of Health, Bethesda, 1981.
21. Gottlieb, M. S., Shear, C. L., and Seale, D. B. Lung cancer mortality and residential proximity to industry. Environ. Health Perspect. 45: 157–164 (1982).
22. Bross, I. Misclassification in 2 × 2 tables. Biometrics 10: 478–486 (1954).

23. Henderson, B. E., Gordon, R. J., Menck, H., Soohoo, J., Martin, S. P., and Pike, M. C. Lung Cancer and air pollution in South Central Los Angeles County. Am. J. Epidemiol. 101: 477–488 (1975).
24. Pike, M. C., Jing, J. S., Rosario, I. P., Henderson, B. E., and Menck, H. R. Occupation: "Explanation" of an apparent air pollution related localized excess of lung cancer in Los Angeles County. In: Energy and Health (N. Breslow and A. Whittemore, Eds.), Proceedings of SIAM-SIMS Conference, Alta, Utah, June 26–30, 1978. SIAM, Philadelphia, 1979, pp. 3–16. 1979.
25. Wesolowski, J. J., Flessel, C. P., Twiss, S., Cheng, J., Chan, R., Garcia, L., Ondo, J., Fong, A., and Lum, S. The chemical and biochemical characterization of particulate matter as part of an epidemiological cancer study. J. Aerosol Sci. 12: 208–212 (1981).
26. Schlesselman, J. J. Assessing effects of confounding variables. Am. J. Epidemiol 100: 3–9 (1978).