Some Effects of Cigarette Smoking, Arsenic, and SO_2 on Mortality Among US Copper Smelter Workers

Philip E. Enterline, PhD; Gary M. Marsh, PhD; Nurtan A. Esmen, PhD; Vivian L. Henderson, MS; Catherine M. Callahan, MS; and Myunghee Paik, MS

Previous studies of the relationship between arsenic levels and respiratory cancer among copper smelter workers have not directly accounted for possible effects of SO_2 exposure and cigarette smoking. This is a report on the 1949–1980 mortality experience of 6,078 white male workers who worked at least 3 years between 1 January 1946 and 31 December 1976 at one or more of eight US copper smelters. The completeness of the cohort was verified statistically, and worker exposures to arsenic, SO_2, dust, nickel, cadmium, and lead were estimated from retrospective industrial hygiene surveys reported elsewhere. By using internal controls, a dose-response relationship for lung cancer was observed with exposure to arsenic and SO_2. When cigarette smoking data were included with arsenic and SO_2 exposure data in a nested case-control analysis, only smoking and arsenic were statistically significant factors. The arsenic-lung cancer relationship was confined to a single smelter associated with high content feed. In the remaining smelters mortality for all causes of death and for all cancer was not high based on comparisons with national, state, and local rates.

There were in recent years 16 copper smelters operating in the United States. Of these, two with high airborne arsenic levels—one in Tacoma, WA and one in Anaconda, MT—have been the subject of epidemiologic investigations in which levels of air arsenic were related to deaths from respiratory cancer.[1–3] A third smelter in Salt Lake City, UT has been studied, but levels of air arsenic were not reported.[4] Studies of all three smelters revealed an excess in respiratory cancer. For the two in which levels of air arsenic were estimated this excess was directly related to arsenic exposure.

This is a report of an historical prospective mortality study and a case-cohort study of workers from eight copper smelters in the United States including the smelter studied earlier for which levels of air arsenic were not estimated. Males who worked a cumulative 3 years or more in copper smelting during the years 1946–1976 were observed for deaths through 1980. Smelters were selected for study because they had been in operation for a long time and had adequate employee records on which to develop a cohort.

Previous studies of copper smelters that show a relationship between arsenic levels and respiratory cancer did not directly account for possible effects of SO_2 exposure and cigarette smoking on this relationship. SO_2 has been reported to be an animal carcinogen, and its irritant properties have been considered as a possible mechanism for a cocarcinogenic or cancer promoting effect.[5,6] Since levels of SO_2 and arsenic often vary together in the environment of smelters, a role for SO_2 in the production of lung cancer among arsenic-exposed workers has not been eliminated. Cigarette smoking is a potential confounder in any study of respiratory cancer.

High exposures to SO_2 are common to nearly all smelting operations since most ores smelted are sulfide ores and the smelting process involves the conversion of these sulfides to oxides. However, arsenic exposure varies depending mainly on the arsenic content of the feed. It was hoped that by studying several smelters with varying levels of arsenic in the feed, contributions made by arsenic to respiratory cancer could be separated from any contributions made by SO_2. Also, by

From the Departments of Biostatistics (Dr Enterline, Dr Marsh, Ms Henderson, Ms Callahan, and Ms Paik) and Industrial and Environmental Health Sciences (Dr. Esmen).

Address correspondence to Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261 (Dr Enterline).

0096-1736/87/2910-0831$02.00/0

adding a nested case-cohort study, any confounding effects of cigarette smoking could also be accounted for.

Methods

The primary data collection phase of this study was conducted between August 1981 and October 1982. During this time a University of Pittsburgh data collection team traveled to each plant and screened personnel, payroll, and medical records for all male production employees and ex-employees who met the study eligibility criteria. All pertinent demographic and work history information was then microfilmed so that time consuming data abstraction activities could be conducted off site. The completeness of the study population as assembled from plant records was verified using the method developed by Marsh and Enterline.[7] The verification scheme involved cross-checking a self-weighted random sample of names that appeared on two or more triannually successive first quarter returns filed by plant management with the Social Security Administration against personnel records abstracted from plant files. Any names appearing on the social security forms (form 941) but not appearing in the cohort were checked against plant records to determine the reason for exclusion. For the most part names on form 941 not in the cohort were ineligible for the study (eg, female, worked <3 years, or no potential for plant exposure such as cafeteria worker, secretary, etc). However, for some, records were in the plant files but missed during the initial survey, and for some, the employee record could not be found. The number of sample records that were in the plant files but had been missed divided by the total sample size adjusted for those missed for acceptable reasons forms the basis for a "missed rate." All plants had "missed rates" below 5%, and this was considered acceptable. However, some names on the social security forms could not be located in plant files. The number of such names divided by the number in the sample is called the "undetermined rate." This was well below 5% for all plants except one (Plant 4). For this plant the rate was 24.6%. It was subsequently found that records from this plant had been transferred to a different location for the purpose of an epidemiologic study. All these records were recovered and added to the study cohort bringing the "undetermined rate" for that plant well below 5%.

One special feature of this study is that for one or more of the smelters it was possible for one of us to estimate exposures to arsenic, SO_2, dust, nickel, cadmium, and lead by job and by year, thus enabling workers to be assigned an exposure level to these contaminants. By utilizing the strategy developed by Marsh,[8] the diverse occupational histories and multiagent exposure data from the various smelters were merged into a uniform format suitable for statistical analysis. Several exposure-disease relationships could be examined as the result of this format and are reported elsewhere.[9] This report summarizes the results of the study only as they relate to exposures to arsenic and SO_2 and to mortality of white males for all causes, all cancer, and lung cancer. There were no females identified as eligible for study and only a small number of blacks, too few to make separate analysis worthwhile.

The total and cause-specific mortality experiences of the white male copper smelter workers were examined during 1 January 1949 to 31 December 1980. All primary analyses were conducted utilizing the Occupational Cohort Morality Analysis Program developed by Marsh and Preininger[10] and later extended by Caplan et al[11] and Marsh et al.[12]

All cause mortality reflects the general levels of health of copper smelter workers, whereas cancer may be more directly related to contaminants in the working environment of copper smelters. Six of the eight copper smelters in this study had a mine and mill operating in conjunction with the smelter, and one of these also had a refinery. Only the mortality experience of workers whose only work experience was at the copper smelter is reported here.

Results

Table 1 shows the copper smelters studied along with some information about each smelter. The oldest of these had a start-up date of 1901, and the newest had a start-up date of 1950. Six of the eight smelters had start-up operations in 1940 or earlier. Although the age of the smelters complicates attempts to calculate exposures it ensures that exposures occurred in the sufficiently distant past and for sufficiently long periods to permit meaningful observations with regard to cancer. Estimated arsenic concentrations of the feed for seven of the eight smelters are also shown by time. Feed includes the processing of spiess as well as copper ore. Plant 3 is the smelter previously studied by Rencher et al[4] and before 1959 operated as a custom smelter reprocessing some materials high in arsenic content from other smelters. Clearly the arsenic content of the feed for Plant 3 is different from the other smelters in this study. Table 1 also shows the number of workers studied, person years studied, mean years worked, the mean level of cumulative and average arsenic exposure, and mean years worked in jobs where SO_2 exposure peaks exceeded 6 and 3 ppm.

A total of 6,078 white male workers were identified as having worked a cumulative 3 years or more between January 1946 and December 1976—ranging from 189 for Plant 5 to 2,288 for Plant 3. As suggested by the arsenic content of the feed, Plant 3 had a considerably higher mean worker arsenic exposure than the other copper smelters studied. For Plant 2 levels of arsenic were too low to warrant estimates of individual worker exposure. For Plant 5 no data were available to make individual exposure estimates for arsenic or any other substance.

Plants also differed somewhat in terms of exposure to SO_2. It was felt that SO_2 exposures cannot be meaningfully expressed as averages, as can arsenic exposure, so that for this analysis worker exposure to SO_2 has been

TABLE 1
Selected Data for Eight US Copper Smelters

Plant No.	State	Start-up Date	% As in Feed (1950)	No. of Workers Studied	Person Yr Studied	Mean Yr Employed	Mean Level of Arsenic ($\mu g/m^3$)	Mean Time-Weighted Arsenic ($\mu g/m^3$-yr)	Mean Years Exposed to SO_2 in Jobs with Peak SO_2	
									>6 ppm	>3 ppm
1	AZ	1912	.02	563	9,016	16.7	12.9	177.9	6.1	10.4
2	TN	1901	<.01	497	10,119	21.6			7.0	11.2
3	UT	1906	.63	2,288	46,347	21.1	68.6	959.7	8.0	10.2
4	NV	1910	.05	602	11,474	18.1	12.0	176.2	0.0	4.4
5	AZ	1924		189	3,492	15.4				
6	AZ	1950	.08	599	11,747	16.3	8.1	124.5	3.7	6.0
7	AZ	1904	.035	965	19,891	21.5	7.0	114.5	4.4	7.9
8	AZ	1940	.03	375	6,518	15.7	7.5	133.4	4.2	4.2

expressed as years in jobs where peak exposures exceeded various levels. For Table 1, 6 and 3 ppm were chosen somewhat arbitrarily to illustrate differences among smelters.

The mortality experience of workers meeting the study criteria was determined for the period 1949 through 1980 by first examining plant records to establish whether these workers were still employed or death was known to the company. The remaining worker records were submitted to the Social Security Administration (SSA), and it was determined whether individuals were alive as of 31 December 1980 as evidenced by the fact that they were still contributing to or were receiving benefits from the SSA. Deaths known to the SSA were also identified. For the remaining workers who were neither known to be dead nor contributing to or receiving benefits from the SSA, clearance was made with state driver's license bureaus, the Veterans Administration, the US Postal Service, and ultimately direct telephone contacts.

Of the total 6,078 white male workers with experience only in copper smelting 73.9% were alive at the end of the study, 24.5% dead, and the status of 1.6% unknown. For those known to be dead, death certificates were located for 94.1%. All deaths were coded by a nosologist to the revision of the International Classification of Diseases in effect at the time of death to facilitate comparisons with official vital statistics tabulations. For this report, cause of death groupings are as defined in the eighth revision of the international lists of causes of death.

Table 2 shows standardized mortality ratios (SMR) for all causes of death, all cancers, and lung cancer where the expected numbers of deaths are based on the national mortality experience for white males, the state mortality experience for white males, and the mortality experience of counties in the area in which the smelter was located, also based on white males. SMR is the ratio of observed to expected number of deaths (×100) for persons of the same age, sex, and race and living during the same period, where the expected is based on some reference population. In this case, the reference population is the total national population, the population of the state in which the smelter is located, and the population of the county or group of counties in the area in which the smelter is located. Whereas for Table 2 it was possible to examine mortality for cancer for the years 1949–1980, mortality for all causes could be examined only for the years 1960–1980 since detailed data by county are available only for those periods. This type of analysis is made possible by the mortality and population data system maintained by the Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh. This system was developed primarily from the National Center for Health Statistics data tapes furnished to the Department of Biostatistics that contain detailed demographic and cause of death information for cancer deaths starting in 1949 and for all deaths starting in 1960. The MPDS is periodically updated through a cooperative arrangement with the Environmental Protection Agency Office of Biometry in Research Triangle Park, NC. That office has also provided

TABLE 2
Observed Deaths and SMRs for Selected Causes of Death

Plant No.	Observed No. of Deaths	SMRs Based on Expected		
		National	State	Local Area
All causes (1960–1980)				
1	93	97.7	101.1	101.2
2	92	96.5	88.8	89.7
3	491	88.8†	101.6	101.6
4	118	78.0†	74.8†	70.8†
5	52	128.9	134.2*	124.3
6	144	94.2	98.2	98.5
7	204	74.5†	78.2†	78.4†
8	41	87.3	87.2	93.7
All cancer (1949–1980)				
1	22	98.2	109.9	107.9
2	20	84.9	89.5	95.0
3	128	98.2	137.6†	132.1†
4	21	56.6†	58.8*	68.5
5	10	94.6	106.4	108.7
6	32	87.6	98.3	96.2
7	41	63.4†	71.0*	74.9
8	3	27.5*	30.4*	35.0
Lung cancer (1949–1980)				
1	8	118.1	127.5	125.0
2	7	90.3	86.3	89.4
3	48	119.7	226.9†	210.9†
4	8	73.9	71.4	97.6
5	4	141.1	149.9	160.8
6	8	70.3	76.0	74.1
7	10	51.4*	55.3	60.5
8	0	0.0*	0.0	0.0

* $P < .05$.
† $P < .01$.

population data needed for calculation of the rates used for calculating expected numbers of deaths.

For most plants the all causes SMRs based on national data approximated the SMRs based on state or county mortality experience. For Plant 3, which is located near Salt Lake City, the use of US mortality data produced an all cause SMR considerably lower than mortality data for Utah and the area around Salt Lake City. For Plant 5 the all cause SMRs were high, and the SMR was statistically significant when a comparison was made with death rates for the state where the smelter is located. This is a plant that closed in 1972 and is the plant for which no exposure information was available.

For all cancer and for lung cancer the SMRs based on national mortality were fairly close to SMRs based on state or local data, except for Plant 3. For Plant 8 the cancer mortality experience is remarkable. At this time we have no explanation for the extremely low observed cancer mortality for this plant aside from the possibility that this could be a chance occurrence. There was nothing learned in the verification process that suggests a reason for the low cancer mortality for workers from this plant. A total of 196 names were sampled from forms 941, as described above. Of these 139 were not in the cohort. However, for 135 of these the reason was acceptable (i.e., females, worked <3 years, etc). For the four with unacceptable reason three were found in the plant files but had been missed at the time of the initial microfilming. The missed rate for this plant was 4.9% whereas the undetermined rate was .5%, using definitions given above. Note, however, that only a sample of names appearing on forms 941 were actually used in the verification process, and it would be easy to miss a few workers who died of cancer since they would constitute only a small part of the population. Thus, the verification procedure is useful primarily for the identification of the entire sets of records that are missing or were missed during microfilming, as was the case in Plant 4 as described above. Only a check of all names appearing on forms 941 with a complete follow-up for missing names would be likely to establish the reason for the low cancer rate in Plant 8 if, in fact, this was other than a chance occurrence.

Because of the variation in the background mortality rates, particularly those for cancer, in the areas where the copper smelters were located and because copper smelter workers may not have a mortality experience like others in the areas in which they live, it is unclear what external reference population should be used. As an alternative to making further external comparisons we have chosen a method for analysis of exposures in relation to mortality using the study population as an internal control. The procedures used are those originally described by Mantel and Haenszel[13] and Mantel[14] and later adapted for use in cohort studies by Hakulinen.[15] The computing method, which has been incorporated into the Occupational Cohort Mortality Analysis Program system,[12] was modeled after those developed by Gilbert and Buchanan.[16] These computations utilized person years by age, calendar time, time since first exposure (latency), and exposure level. Expected number of deaths in each exposure level were calculated by multiplying the number of person years in that exposure level and stratum (age, calendar time, latency, etc) by the death rate for the combined exposure categories in that stratum. For example, if there are a thousand person years and ten deaths from a cause of interest in a particular stratum and if there are 400, 300, 200, and 100 person years, respectively, in each of four exposure categories within that stratum the number of expected deaths in these categories would be 4, 3, 2, and 1, respectively. The total number of deaths expected in a given exposure category is obtained by summing over strata. For arsenic exposure, person years at risk relative to exposure categories were lagged for 5 years since it was not believed that arsenic would produce any cancers in < 5 years. However, for SO_2 we are uncertain what (if anything) is produced, and no lag was introduced.

For each exposure category a summary measure was calculated as a Mantel-Haenszel relative risk. This can be thought of as risk in a particular exposure category relative to the risk for pooled exposure categories summed across strata. A test of whether the number of deaths observed in a particular exposure category differs significantly from that expected is the Mantel-Haenszel test statistic as adapted by Hakulinen,[15] and this was used for assessing overall differences in the relative risk for the various exposure categories. In addition, Mantel's test for trend[14] was used to assess whether the relative risks increase significantly with increasing exposure. For this test an exposure value or score was assigned to each exposure category, and observed and expected means of these scores (obtained from weighting the observed and expected deaths, respectively) were calculated and compared. The variance of the difference in the observed and expected scores was then calculated to obtain an approximately normally distributed test statistic.

Table 3 shows the results of an analysis of deaths from lung cancer in relation to individual arsenic exposure expressed as a time-weighted exposure. That is, air arsenic levels in each job and period ($\mu g/m^3$) times the amount of time workers spent in those jobs (years), summed across all jobs and time intervals. Only Plants 1, 3, 4, 6, 7, and 8 are used since no exposure data were available for Plant 5, and for Plant 2 the arsenic content of the ore was so low that estimates of individual exposure levels were not attempted. Also, only those men whose work histories indicated some exposure to arsenic were included in this analysis.

Table 3 shows that the hypothesis of no difference among exposure categories could not be rejected with a two-sided probability of .28. However, the exposure score for the trend test had a P value of .059, which approaches statistical significance. A trend in relative risks is very clear, rising from .58 to 1.60, and when plotted on an arithmetic scale is concave downward.

Exposures to arsenic reported in the six copper smelters represented by Table 3 were low relative to exposures at two copper smelters studied earlier and not included here.[2,17] For example, the smelter with the

TABLE 3

Observed and Expected* Deaths from Lung Cancer and Relative Risks by Cumulative Arsenic Exposure for Copper Smelter Workers Exposed to Arsenic (Lagged 5 yr)

	Arsenic Time-Weighted Exposure ($\mu g/m^3$ yr)				Test of Heterogeneity†	Exposure Score for Trend Test		Trend Test P Value§
	<100	100–249	250–999	1,000+		Observed	Expected	
Observed	12	18	22	18				
Expected	17.99	20.19	19.30	12.51	.28	892.86	682.52	.059
Relative Risk	0.58	0.85	1.21	1.60				

* Expected deaths based on experience of all white copper smelter (only) workers adjusting for age, calendar year, and latency period.
† Two-sided probability based on three degrees of freedom χ^2 test.
§ One-sided probability of trend arising due to chance.

TABLE 4

Observed and Expected* Deaths from Lung Cancer and Relative Risks by Duration of Exposure to SO_2 Greater than or Equal to Selected Peak Values for Copper Smelter Workers Exposed to SO_2

Peak SO_2 (ppm)	Duration (yr)				Test of Heterogeneity†	Exposure Score for Trend Test		Trend Test P Value§
	<5	5–9	10–14	15+		Observed	Expected	
0.75								
Observed	14	10	11	43				
Expected	15.5	11.0	9.2	42.3	.95	14.20	13.87	.46
Relative risk	0.88	0.89	1.24	1.04				
3.00								
Observed	21	9	9	25				
Expected	21.4	11.5	7.5	23.6	.95	11.45	11.03	.40
Relative risk	0.97	0.75	1.24	1.10				
6.00								
Observed	18	3	7	17				
Expected	20.9	6.1	5.6	12.4	.58	11.00	9.24	.13
Relative risk	0.76	0.44	1.31	1.62				
12.00								
Observed	10	2	6	3				
Expected	13.1	2.6	1.6	3.7	.97	8.33	6.96	.03
Relative risk	0.55	0.73	5.21	0.76				
24.00								
Observed	2	2	3	0				
Expected	4.4	1.1	0.8	0.8	.98	8.21	6.41	.01
Relative risk	0.23	2.27	5.59	0.00				

* Expected deaths based on experience of all white copper smelter only workers adjusting for age, calendar year, and latency period.
† Two-sided probability based on three degrees of freedom χ^2 test.
§ One-sided probability of trend arising due to chance.

highest exposure (Plant 3) had a mean worker exposure of 68.6 $\mu g/m^3$ and a mean time-weighted level of 957.7 $\mu g/m^3$-yr. In a recent reanalysis of data from the copper smelter at Tacoma, WA the mean worker exposure was 520 $\mu g/m^3$, and the mean time-weighted exposure was 6,530 $\mu g/m^3$-yr.[17] The highest exposure category shown in Table 3 roughly corresponds to the two lowest exposure categories in the recent Tacoma report. A recent report on the Anaconda smelter yields exposure levels similar to those for Tacoma. The mean worker exposure was 685 $\mu g/m^3$, and the mean time-weighted exposure was 10,400 $\mu g/m^3$-yr.[2]

Table 4 is similar to Table 3 but deals with respiratory system cancer in relation to SO_2. Only those men whose work histories indicated some exposure to SO_2 are included. As in Table 3 none of the tests for heterogeneity are statistically significant; however, the trend test is statistically significant ($P = .03$) at SO_2 peak exposure levels exceeding 12 ppm.

To examine the separate effects of arsenic and SO_2, giving consideration to cigarette smoking, a case-cohort study was nested in the study already described. All copper smelter workers who died of lung cancer between 1 January 1949 and 31 December 1980 were identified as cases. Controls were a random sample of all workers with stratification by plant and year of birth. In selecting controls certain members of the original cohort were considered ineligible. Excluded were members of the cohort who died of lung cancer, those dying of any cause before 1 January 1949, those with unknown date of death, those dying before age 45, those <45 years old as of the termination of the study (31 December 1980), those who terminated employment before 1 January 1949, those whose year of termination of follow-up minus year of birth was <45, those with unknown date of termination, and those with unknown date of birth. Controls represented a 5% stratified random sample given restrictions outlined above. A case-cohort study design has been described by Prentice et al[18] and was chosen partly because it allowed interviewing for smoking histories to start before lung cancer cases were known and because it yields self-weighted estimates of smoking for workers from each smelter studied.

Smoking histories were obtained through telephone

interviews with the worker (if still living) or a knowledgeable informant, ideally a member of the worker's immediate family. Smoking data were obtained for 76% of the lung cancer cases and for 85% of the control group. The difference was probably because most controls were alive as of the date the study was conducted, whereas all cases were deceased and it was more difficult to obtain information on dead individuals than living individuals. When cases were compared with controls who had died before the end of the study, completion rates were quite similar.

Cases and controls for whom smoking data were obtained were compared with those from whom smoking data was not obtained on a large number of variables related to both lung cancer and probable occupational exposure levels, and it was concluded that the discrepancy in noncompletion rates between cases and controls has not introduced any significant bias. Data were analyzed by logistic regression with backward stepwise elimination using the statistical package BMDP.[19] There were two basic areas of interest—matching or confounding variables, including cigarette smoking and demographic variables, and occupational exposures. Cases and controls were initially analyzed to identify those smoking and demographic variables in combinations that best explained lung cancer mortality. The smoking variable chosen was a combination of duration of smoking and time since smoking began, whereas the only important demographic variable was age at death or age on 31 December 1980.

There were 183 copper smelter workers for whom data regarding smoking, arsenic exposure, and SO_2 exposure were available. These workers were from Plants 1, 3, 4, 6, 7, and 8, represented in Tables 3 and 4. As in Table 3, arsenic was expressed as a time-weighted measure. SO_2 exposure was somewhat arbitrarily expressed as duration of exposure to SO_2 in jobs where peaks exceeded 6 ppm. Other exposure levels were examined, and results were generally supportive of the results at 6 ppm. An examination of the data set revealed that for 181 of 183 copper smelter workers, arsenic exposure levels ranged from zero to 4,196 $\mu g/m^3$-yr, whereas two workers had exposures of 9,982 and 10,503, respectively, one a case and one a control. It was decided that these two outliers could distort the analysis, and they were dropped.

The results of a logistic regression analysis performed on copper smelter workers are shown in Table 5. All major variables shown in Table 5 plus age were entered into the regression analysis, along with all four-way, three-way and two-way interactions among exposure and smoking variables (16 terms). Table 5 shows terms that were statistically significant ($P < .05$) for smelter workers in the case-cohort study and smelter workers from Plant 3. Data in original units and square roots of the original units were used. The fit of the model is shown by the P values at the bottom of each column.

For data in original units the logistic coefficients for smoking and arsenic exposure were statistically significant and not much different for workers from all plants and workers from Plant 3. When only workers from plants other than Plant 3 were used the only statistically significant variables were those related to smoking (not shown). Thus the relationship between arsenic and lung cancer is apparently driven by Plant 3. Since cases and controls were unmatched, some additional matching variables (original units) were added to those in Table 5. Only Plant 3 was used here since all information about arsenic and lung cancer appears to be contained in this data set. Matching variables entered together were year of birth and year of hire. These had little effect on the coefficients shown in Table 5. The coefficient for arsenic, for example, was changed from .000816 to .000921. Also, removal of smoking variables had little effect on the coefficient for arsenic.

When original units were used, the coefficient for arsenic exposure was surprisingly large, predicting unrealistically high relative risks when arsenic exposures like those for the smelter at Tacoma, WA were entered. Moreover, as noted for Table 3, when relative risks were plotted against dose the dose-response curve was concave downward, like that reported for the Tacoma, WA and Anaconda, MT smelters, whereas the predicted dose-response curve from a logistic regression for exposures in the range of interest is concave upward.[17]

To alter the shape of the dose-response curve to give more realistic projections at higher levels of arsenic exposure, data were transformed by taking square roots. A shown in Table 5, when this was done only years smoked and arsenic exposure were statistically significant for the full data set. When analysis was confined to workers from Plant 3, some two-way interaction terms were statistically significant. This transform gives a dose-response curve for arsenic that is slightly concave downward and gives fairly realistic response estimates at extrapolated higher arsenic exposure levels.

Because the controls were actually a random sample of the eligible population, a profile of smoking habits for copper smelter workers may be obtained. There were 126 controls for whom smoking histories were obtained. Of these, 70.6% gave a history of cigarette smoking. For Plant 3 this percentage was 64.7, while for the remaining plants taken together this percentage was 74.7 The difference in smoking habits between Plant 3,

TABLE 5

Statistically significant log odds ratios for lung cancer with backward elimination after entry of years smoked, years since started smoking, time-weighted exposure to arsenic, years in job with SO_2 exposure greater than 6 ppm, all four-way, three-way, and two-way interactions, and age

Variables	Original Units (x)		Transformed Units ($\sqrt{x}$)	
	All workers (181)	Plant 3 workers (86)	All workers (181)	Plant 3 workers (86)
Years smoked (x1)	0.0799	0.114	0.199	1.27
Years since started (x2)	0.0197	0.0203	NS	−0.0294
Interaction (x1, x2)	−0.00132	−0.00221	NS	−0.15
Arsenic (x3)	0.00104	0.000816	0.0439	0.0093
SO_2 (x4)	NS*	NS	NS	−0.697
Interaction (x3, x4)	NS	NS	NS	0.0192
P (goodness of fit)	.642	.666	.343	.474

* NS not significant at 5% level.

located in Utah, and all other plants was not as large as might have been expected in view of the fact that the lung cancer death rate for Utah is only about half the rate for the rest of the United States, and the difference is probably due to cigarette smoking. This suggests that smelter workers at Utah were not typical of the Utah population with regard to smoking behavior and that SMRs for lung cancer using Utah rates to calculate expected deaths are probably overstated.

Discussion

This report examines a part of a larger data set generated as a result of examining the mortality experience of copper smelter workers in the United States. It shows that all cause mortality rates in copper smelter workers are not unusually high and for lung cancer the one smelter that appeared to have a problem was the one with an appreciable exposure to arsenic. Arsenic clearly emerged as a primary cause of lung cancer in smelter workers after adjusting for cigarette smoking and SO_2 exposure.

An important question is whether there is an interaction among smoking, arsenic, and SO_2. The logistic model used for the case-cohort study is a multiplicative model so that lack of interaction in a logistic model specifies lack of interaction other than multiplicative interaction. Since no significant interactions could be detected when data were analyzed in their original form it can be inferred that interactions examined did not deviate from simple multiplicative interactions. What we can say, therefore, from the case-cohort analysis is that we are unable to reject the notion that arsenic, SO_2, and smoking have multiplicative effects. However, we did not test for multiplicative effects so that some other form of analysis is needed to provide more direct information on this subject.

This study illustrates the difficulty in finding a suitable population for calculating expected numbers of deaths in an occupational cohort. For the copper smelter located in Salt Lake City, UT where there are large numbers of Mormons who abstain from cigarette smoking, calculations of standardized mortality ratios for lung cancer using the US mortality experience as the standard seemed inappropriate. Moreover, workers at the Salt Lake City plant did not appear to be representative of the population of the area in which the smelter is located. Thus, the true SMR for lung cancer for the Utah smelter workers apparently lies between 119.7 (US expected) and 226.9 (Utah expected). A rough estimate of this can be obtained by calculating the excess in respiratory cancer predicted at the Utah smelter given data on arsenic levels at that smelter and knowledge of the relationship between arsenic and lung cancer from studies of other smelters. A recent reanalysis of data from the copper smelter located in Tacoma, WA shows that a power function provided a good fit of the dose-response data available for that smelter.[17] This was confirmed when a similar data set was reanalyzed from a smelter at Anaconda, MT. The average time-weighted exposure to arsenic for workers at the Salt Lake City smelter was 959.7 $\mu g/m^3$-yr. By using the power function derived from the Tacoma data, this predicts a standardized mortality ratio 154.0, an SMR that falls between the SMR off 119.7 calculated for the Utah smelter using the US white male population to calculate expected numbers of deaths and the SMR of 226.9 using the Utah population to calculate expected deaths.

This study also suggests that if there are a few missing records in an occupational cohort study their absence is extremely difficult to detect using methods proposed to date.[7] For one smelter in this study the death rate for cancer was very low, and this could have been due to missing records for workers who died of that disease. Although this smelter does not greatly influence overall results, the fact that such situations might exist is disturbing and detracts from the validity of occupational epidemiology.

Finally, this study represents an experience in applying logistic regression analysis to occupational data and demonstrates the need to actually test the resulting model not only by goodness-of-fit tests but also by seeing if it has believable predicting ability. Since the logistic model is multiplicative and the dose-response curve concave upward, extrapolation to arsenic levels much above those from which the model coefficients were derived resulted in extraordinarily high predicted risks for lung cancer. An attempt was made to correct for this by a transformation of variables. However, what transformations to make and what variables to transform is a difficult decision. Clearly there is no reason to believe that all independent variables have the same relationship to the dependent variable nor with each other. Care must be exercised therefore in extrapolating coefficients developed from logistic or any other form of regression analysis.

This is a study of the mortality experience of workers from eight US copper smelters. For each copper smelter an attempt was made to estimate worker exposure for various contaminants. Presented here is the mortality experience for all causes of death, all cancer, and lung cancer. For lung cancer there was a dose-response relationship with exposure to arsenic and with exposure to SO_2 using an internally generated control group. When data on cigarette smoking were added to exposure data for arsenic and SO_2 in a nested case-cohort analysis only smoking and arsenic were statistically significant factors. The relationship between arsenic and lung cancer was apparently confined to a single smelter where the feed had been relatively high in arsenic content. Arsenic exposure did not appear to be related to lung cancer in the other smelters, and mortality for all causes of death and for all cancer was not high.

Acknowledgments

This research was supported under a contract between the Smelter Environmental Research Association and the Department of Biostatistics, University of Pittsburgh.

The authors greatly acknowledge the assistance and cooperation of Donald Robbins and Gerald Schurtz as well as others in the Smelter Environmental Research Association member companies who participated in this research effort. In addition we acknowledge with great appreciation the staff at the University of Pittsburgh: Michele LaValley, Andrea March, and Donald Fisherowski who collected and coded demographic and work history data and conducted follow-up and verification, computer programmers Mary Preininger, Marcia Goetsch, and Jay Graham, Cathy Champagne who collected the data on tobacco smoking, and student assistants Ella Lyons who participated in the statistical analysis, Donald Tuchman who assisted in collecting and evaluating industrial hygiene data, and Gloria Claus and Linda Kobus who patiently typed and corrected this manuscript.

References

1. Lee AM, Fraumeni JF: Arsenic and respiratory cancer in man: An occupational study. J Natl Cancer Inst 1969;42:1045–1052.

2. Higgins ITT, Welch KB, Oh MS, et al: Arsenic exposure and respiratory cancer in a cohort of 8,044 Anaconda smelter workers. Report to the Chemical Manufacturers Association and the Smelter Environmental Research Association, Apr 1986.

3. Enterline PE, Marsh GM: Cancer among workers exposed to arsenic and other substances in a copper smelter. *Am J Epidemiol* 1982;116:895–911.

4. Rencher AC, Carter MW, McKee DW: A retrospective epidemiological study of mortality at a large copper smelter. *J Occup Med* 1977;19:754–758.

5. Peacock PR, Spence JB: Incidence of lung tumors in LX mice exposed to (1) free radicals; (2) SO_2. *Br J Cancer* 1967;21:606–618.

6. Amdur MO: Effects of sulfur oxides on animals, in Nriagu NO (ed): *Sulfur in the Environment, Part II: Ecological Impacts*. New York, John Wiley and Sons, 1978, pp 62–74.

7. Marsh GM, Enterline PE: A method for verifying the completeness of cohorts used in occupational mortality studies. *J Occup Med* 1979;21:655–670.

8. Marsh G: A strategy for merging and analyzing work history data in industry-wide occupational epidemiology studies. *Am Ind Hyg Assoc J* 1986;48:414–419.

9. Enterline PE, Marsh GM, Esmen NA, et al: Mortality among copper and zinc smelter workers in the United States. Final Report to the Smelter Environmental Research Association. University of Pittsburgh, Pittsburgh, PA, Jan 1986.

10. Marsh GM, Preininger ME: OCMAP: A user-oriented occupational cohort mortality analysis program. *Am Stat* 1980;34:245.

11. Caplan RJ, Marsh GM, Enterline PE: A generalized effective exposure modelling program for assessing dose-response in epidemiological investigations. *Computers Biomed Res* 1983;16:587–596.

12. Marsh GM, Ehland J, Paik M, et al: OCMAP/PC: A user-oriented occupational cohort mortality analysis program for the IBM/PC. *AM Stat* 1986;40:308.

13. Mantel N, Haenszel W: Statistical aspects of the analysis of data from retrospective studies of disease. JNCI 22(1958)719–748.

14. Mantel N: Chi-square tests with one degree of freedom: Extensions of the Mantel Haenszel procedure. *J Am Stat Assoc* 1963;58:690–700.

15. Hakulinen T. Mantel-Haenszel statistic for testing the association between a polychotomous exposure and a a rare outcome. *Am J Epidemiol* 1981;113:192–196.

16. Gilbert ES, Buchanan JA: An alternative approach to analyzing occupational mortality data. *J Occup Med* 1984;26:822–828.

17. Enterline PE, Henderson VL, Marsh GM: Exposure to arsenic and respiratory cancer. A reanalysis. *Am J Epidemiol* 1987;125:929–938.

18. Prentice RL, Self SG, Mason MW: Design options for sampling within a cohort in Moolgavkar SH, Prentice RL (eds): *Modern Statistical Methods in Chronic Disease Epidemiology*. New York, John Wiley and Sons, 1986.

19. Dixon WJ: *BMDP Statistical Software*. Berkeley, CA, University of California Press, 1981.