American Journal of Industrial Medicine 21:491–506 (1992)

Adjustment for Smoking, Alcohol Consumption, and Socioeconomic Status in the California Occupational Mortality Study

James J. Beaumont, PhD, James A. Singleton, MS, Gwendolyn Doebbert, MS, Kathryn R. Riedmiller, BA, Robert M. Brackbill, PhD, and Kenneth W. Kizer, MD

This paper presents methods for adjusting for smoking, alcohol, and socioeconomic status in death certificate-based occupational mortality surveillance. The methods were applied in the California Occupational Mortality Study, a statewide study of rates based on 180,000 deaths and census estimates of occupations. For each occupation, levels of smoking, alcohol consumption, and socioeconomic status were estimated using National Health Interveiw Survey and U.S. Census data, and an empirical Bayes procedure was used to improve the stability of smoking and alcohol estimates for small occupations. Expected death rates for occupations were calculated by modeling rates as a function of age, smoking, alcohol, and socioeconomic status with Poisson regression. The effect of adjustment was usually moderate and in the expected direction, and the adjusted mortality ratios were generally closer to 1.0. Full data on agricultural occupations are presented for illustration.

Key words: epidemiologic methods, alcohol use, census data, occupational mortality, surveillance, agricultural occupations, death certificates, blacks, NHIS

INTRODUCTION

Surveillance of occupational mortality using death certificates to identify occupation and cause of death is an important means of identifying occupational risk, particularly in occupations where workers are migratory or are employed in small workplaces where it is difficult to assemble cohorts for study [Melius et al., 1989]. Thirty-one states and the District of Columbia conduct death certificate-based occupational mortality surveillance [Dubrow et al., 1987].

Division of Occupational and Environmental Medicine, Center for Occupational and Environmental Health, University of California, Davis (J.J.B., J.A.S.).
California Department of Health Services, Sacramento (G.D., K.R.R., K.W.K.).
Surveillance Branch, Division of Surveillance, Hazard Evaluations and Field Studies, National Institute for Occupational Safety & Health, Cincinnati, OH (R.M.B.).
K.W.K. is currently employed by the Department of Community Health, University of California, Davis.
J.A.S. is currently employed by the California Department of Health Services, Office of AIDS.
Address reprint requests to James J. Beaumont, Ph.D., Division of Occupational and Environmental Medicine, University of California, Davis, CA 95616.
Accepted for publication August 7, 1991.

Before an excess of disease can be attributed to occupational exposure, the effects of possible confounding factors need to be considered. Smoking, alcohol, and socioeconomic status have generated the most interest as possible confounding factors since they are related to disease and tend to vary between occupations. Death certificates do not contain information on these potential confounders, however, and direct control for their effects is not possible.

In this paper we explore use of National Health Interview Survey and U.S. Census data to indirectly adjust for smoking, alcohol, and socioeconomic status. The occupational mortality data we used was from the California Occupational Mortality Study, a study of 180,000 California death certificates from 1979 to 1981 for persons aged 16 to 64, and 1980 U.S. Census estimates of occupations in California [California Department of Health Services, 1987]. The California Occupational Mortality Study compared age-adjusted rates in each occupation to the rates for all occupations combined.

We report here methods used for further adjustment for smoking, alcohol, and socioeconomic status, and the effects of adjustment on occupational risk estimates. Readers interested in adjusted risk estimates for specific occupations are referred to a State of California report that has detailed occupation and cause of death tables [Singleton and Beaumont, 1989].

MATERIALS AND METHODS

While the actual smoking and alcohol consumption habits and socioeconomic status of the individual decedents were unknown, estimates of these factors for occupations were constructed using data from national surveys. California survey data were available, but sample sizes were too small to produce reliable estimates for individual occupations.

Smoking Data

We obtained smoking data for occupations from tabulations of National Health Interview Survey public use data tapes. In 1978, 1979, and 1980, a smoking supplement was administered to a one-third subsample of persons 17 years of age or older, resulting in a 3-year combined sample of 49,715 people [Brackbill et al., 1988]. For each occupation we calculated an age-specific smoking index by assigning values to categories of smoking as follows: 1 for never-smokers, 3 for former smokers, 5 for current smokers averaging less than 20 cigarettes per day, and 10 for current smokers averaging 20 or more cigarettes per day. These values were multipled by the estimated proportions of people in the smoking categories, and the products were added. For example, an occupation with 40% never-smokers, 20% former smokers, 30% light smokers, and 10% heavy smokers had a smoking index of $(.4 \times 1) + (.2 \times 3) + (.3 \times 5) + (.1 \times 10) = 3.5$.

Alcohol Consumption Data

The data on alcohol consumption came from the H1 Supplement of the National Health Interview Survey of 1977 in which approximately 23,000 persons 20 years of age or more were interviewed [National Center for Health Statistics, 1978; Schoenborn and Danchick, 1980]. The respondents were asked how often they drank (never, occasionally, once or twice a week, or more than twice a week) and how many drinks they usually had at one sitting. We estimated the number of drinks per week for each person in the survey by multiplying the *quantity* (usual number of drinks at one sitting) by the *frequency* (usual number of drinking occasions). For frequency, the value we assigned to each response category was 0 for never drinking, 5 for occasionally, 1.5 for drinking once or twice a week, 4 for more than twice a week, and 8 for eight or more drinks at one sitting. For each occupation we then estimated the age-specific percentage of persons drinking seven or more drinks per week.

Socioeconomic Status

Nam-Powers occupational status scores were used to assign a measure of socioeconomic status to each occupational group [Nam and Powers, 1983]. The scores were based upon education and income estimates in the national 1980 Census, and they ranged from 0 (lowest socioeconomic status) to 100 (highest socioeconomic status). Because the published Nam-Powers scores were for very detailed occupation categories in the U.S. Census, it was necessary to combine the scores for the broader categories used in the California Occupational Mortality Study. For each grouping of occupations we calculated a weighted average of the Nam-Powers scores for specific occupations, where the weight for a Nam-Powers score was the 1980 California census estimate for the occupation. Because the Nam-Powers scores were not age-, race-, or sex-specific, we used the same score for subcategories of age, race, and sex in the regression analyses.

Calculation of Final Smoking and Alcohol Consumption Scores

The reliability of the smoking and alcohol consumption estimates varied considerably among occupations due to differences in sample sizes in the national surveys. The sample sizes for smoking among white males, for example, ranged from 12 in water transportation workers to 2,215 in managers. To minimize inaccuracy due to small survey sample sizes for uncommon occupations, we employed an empirical Bayes procedure to shrink unstable estimates toward the mean for all occupations combined [Efron and Morris, 1977; Morris, 1983]. Empirical Bayes procedures have previously been used in epidemiologic research for geographic comparisons where some rates are unstable due to small population sizes [Efron and Morris, 1975; Manton et al., 1987; Williams et al., 1980].

The Bayes-adjusted scores for smoking and alcohol were a weighted average of the preliminary scores and the scores of all occupations combined. The final score for each occupation was of the form

$$\text{Score}_{\text{final}} = (1-b)\,(\text{Score}_{\text{preliminary}}) + (b)\,(\text{Score}_{\text{all occupations}})$$

where the shrinking factor b was a number between zero and one. The shrinking factor was a function of the sampling variation of the occupation-specific score and the variation of scores between occupations [Morris, 1983]. As the number of respondents per occupation became smaller (e.g., the sampling variability increased), the shrinking factor increased and the final score was closer to the score of all occupations combined. This procedure, described in more detail in the Technical Appendix, was applied separately for smoking and drinking for the four gender/race subgroups. For illustration, Table I shows survey sample sizes, preliminary scores, variances, shrinkage factors, and final scores for alcohol consumption in agricultural

TABLE I. Example of Adjustment of Estimated Levels Alcohol Consumption Based on Sampling Variability in the National Health Interview Survey*

Occupation	N[a]	United States population	Alcohol score (% consuming ≥ 1 drink per day) Preliminary score	Variance[b]	Shrinking factor[c]	Final score
Farmers	147	870,395	15	0.00087	0.16	17
Farm workers	49	303,689	14	0.00240	0.35	19
Gardeners and other agricultural workers	35	225,036	27	0.00566	0.55	29

*Shown for agricultural occupations for white males are the preliminary alcohol consumption scores, survey sample sizes, variances, shrinkage factors, and final scores. The adjustment was performed for alcohol and smoking in the California Occupational Mortality Study.
[a]No. of survey respondents in the National Health Interview Survey 1977 H1 Supplement.
[b]Estimated variance of the preliminary alcohol score.
[c]Empirical Bayes shrinking factor.

occupations. The final scores for all three confounders (smoking, alcohol, and socioeconomic status) are shown in Table II.

Calculation of Expected Rates

Expected rates for occupations were calculated with Poisson regression models that included parameters for age, smoking, alcohol consumption, socioeconomic status, and parameters for two-way interactions between age and the other variables. Poisson regression was used instead of ordinary regression because mortality count data are discrete, non-normal, and have non-constant variability as a function of size [Frome and Checkoway, 1985; Lovett et al., 1986]. Estimation was by the iteratively weighted least squares method using a SAS macro program modified for use under the VMS oprating system (Suissa and Adam, McGill University, personal communication). Because of the large number of models estimated and limited resources available, the same parameters were fit for each cause of death.

The regression models were based upon 268 observed death rates in males (67 occupations by 4 age groups) and 116 observed rates in women (29 occupations by 4 age groups). The age groups were 16–44, 45–54, 55–59, and 60–64. Models were developed for each of the 22 cause of death categories within each of the four gender and race subgroups, for a total of 88 models. For the 22 black female models, only age and socioeconomic status were used, because there was little variability in smoking and alcohol consumption scores between occupations after empirical Bayes adjustment for small survey sample sizes. Male water transportation workers were excluded from analysis because the unadjusted mortality ratios were unusually high and it appeared that there was a major discrepancy between census and death certificate recording of this occupation.

Expected numbers of deaths and mortality ratios were calculated with a computer program designed for analysis of census-based occupational mortality studies [Singleton et al., 1989]. Confidence intervals and probabilities were calculated for the mortality ratios with an exact Poisson test when the observed numbers of deaths were 6 or less, and with the Byar approximation to the exact test when the observed deaths were greater than 6 [Rothman and Boice, 1979].

TABLE II. Summary of Final Scores for Smoking, Alcohol Consumption, and Socioeconomic Status for White Males in Agricultural Occupations in the California Occupational Mortality Study

Occupation	California population	Final scores		
		Smoking[a]	Alcohol[b]	SES[c]
Farmers	38,708	3.6	17	44
Farm workers	120,769	3.9	19	10
Gardeners and other agriculture workers	72,429	4.1	29	18
All occupations	5,522,488	4.5	30	55

[a]Weighted average of 1 for non-smokers, 3 for former smokers, 5 for less than 20 cigarettes per day, and 10 for 20 or more cigarettes per day.
[b]Percent consuming one or more drinks per day.
[c]Nam-Powers socioeconomic status scores (100 = highest).

RESULTS

Figure 1 shows the distribution of the final smoking, drinking, and socioeconomic status scores for the 68 white male occupational groups. Similar plots were constructed for the other race-sex groups. In order to show how death rates varied with levels of the confounders, we predicted rates over a range of values for each potential confounder while holding the other two confounders constant at their median values. Figures 2 through 11 show the predicted rates for lung cancer, cirrhosis of the liver, and all causes of death combined as a function of smoking, alcohol, and socioeconomic status. The rates were strongly positively associated with smoking and alcohol in older age groups, and were strongly negatively associated with socioeconomic status (high SES was predictive of low death rates).

Significance of the Regression Models

Table III shows the significance of models for lung cancer, cirrhosis of the liver, and all causes of death combined. These causes of death were chosen for significance testing because they are known to be related to smoking, alcohol, and socioeconomic status, and our models should predict these causes well if the data and methods are adequate. The probabilities in Table III test the null hypothesis that the full prediction models fit the data no better than the baseline models with age only. This was equivalent to testing that the coefficients of the additional parameters equalled zero. In all cases the full models fit the data significantly better than the baseline models with age only.

Comparison of Original and Confounder Adjusted SMRs

The majority of the significantly high SMRs in the original age-adjusted COMS analysis remained significantly high after further adjustment. Many SMRs became non-significant with further adjustment, however, and a few became significantly high that originally were not. There were a few occupations where adjustment did not have the effect that might be predicted; for example, white female LVNs and health aides had a high mortality ratio for cirrhosis and other liver diseases before adjustment, suggesting possibly heavy drinking. Adjustment for alcohol consumption elevated the mortality ratio further, suggesting an occupational cause for the cirrhosis or

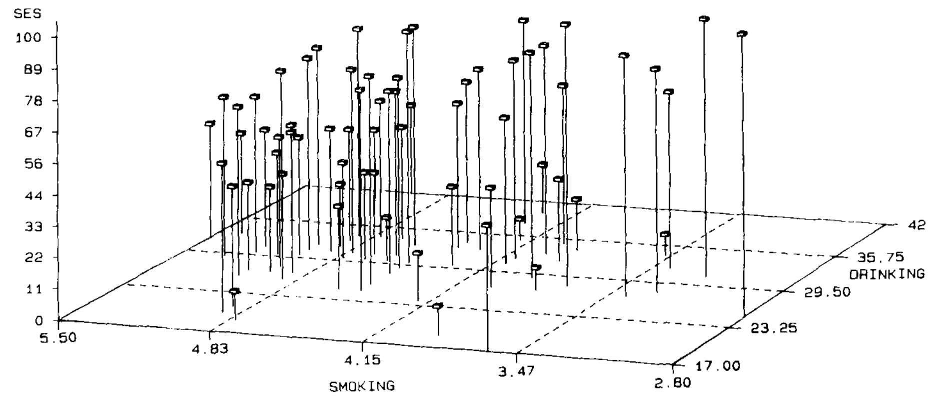

Fig. 1. Distribution of the final smoking, alcohol consumption, and socioeconomic status (SES) scores for the 68 white male occupations in the California Occupational Mortality Study. The smoking score was a weighted average of 1 for non-smokers, 3 for former smokers, 5 for less than 20 cigarettes per day, and 10 for 20 or more cigarettes per day. The alcohol score was the percent consuming one or more drinks per day. The Nam-Powers socioeconomic status score was based on income and education (100 = highest status).

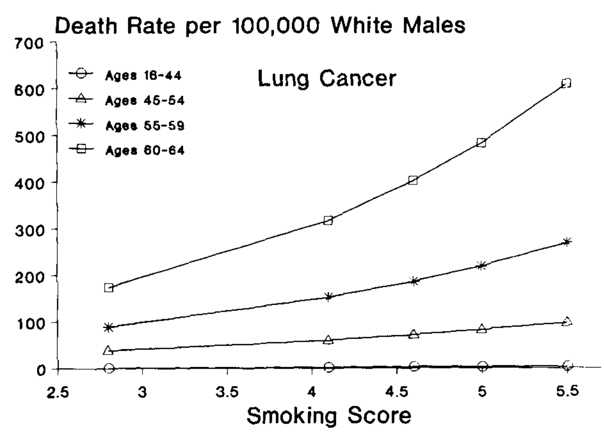

Fig. 2. Predicted death rates for lung cancer by smoking score and age in white males. Rates were annual rates per 100,000 estimated in the California Occupational Mortality Study. Alcohol and socioeconomic status were held constant at their median values. Source of smoking data for occupations: National Health Interview Survey, 1978–1980, Smoking Supplement Subsample.

a problem with use of national data on alcohol consumption for California LVNs and health aides.

To illustrate the effects of adjustment, standardized mortality ratios (SMRs) before and after adjustment for agricultural occupations are shown in Table IV. In farmers, the SMR for all causes of death combined increased from 1.0 to 1.2 with adjustment, a reflection of low smoking and drinking among farmers. Most of the increase in the all-causes SMR was due to increased SMRs for smoking related diseases, in particular heart disease (increased from 1.1 to 1.2) and lung cancer

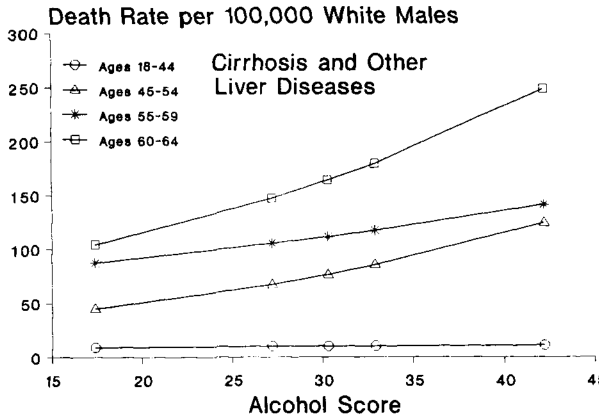

Fig. 3. Predicted death rates for cirrhosis of the liver by alcohol consumption score and age in white males. Rates were annual rates per 100,000 estimated in the California Occupational Mortality Study. Smoking and socioeconomic status were held constant at their median values. Source of alcohol consumption data for occupations: National Health Interview Survey, 1977, H1 Supplement.

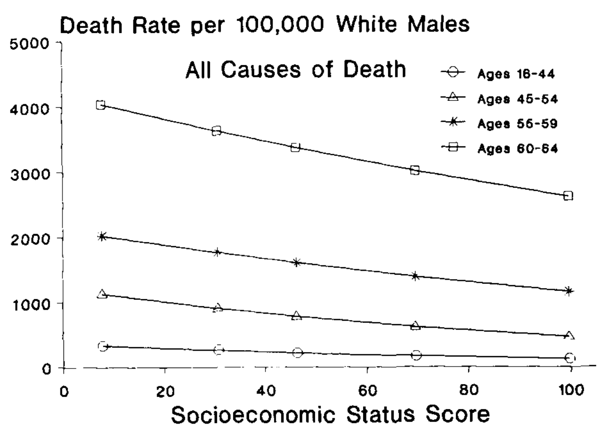

Fig. 4. Predicted death rates for all causes combined by socioeconomic status score and age in white males. Rates were annual rates per 100,000 estimated in the California Occupational Mortality Study. Smoking and alcohol were held constant at their median values. Source of socioeconomic status data for occupations: Nam-Powers Occupational Status Scores, Center for the Study of Population, Florida State University, 1986.

(increased from 0.7 to 1.1). A notable finding after adjustment was significant excess risk of death from chronic obstructive pulmonary disease (SMR 1.7, $p<0.05$).

The smoking and drinking scores for farm workers were well-below average, but those favorable scores were counterbalanced by low socioeconomic status, and the net effect of adjustment was little overall change in expected mortality. After adjustment, the SMR for all causes of death combined was 1.4 ($p<0.05$) due to high risks for accidental injury, cerebrovascular disease, and cirrhosis of the liver. In gardeners and other agricultural workers, the SMR for all causes combined decreased with adjustment from 1.0 to 0.8 due to higher predicted mortality from low socioeconomic status. After adjustment, falls and machinery accidents occurred significantly more often than expected (SMR 1.7).

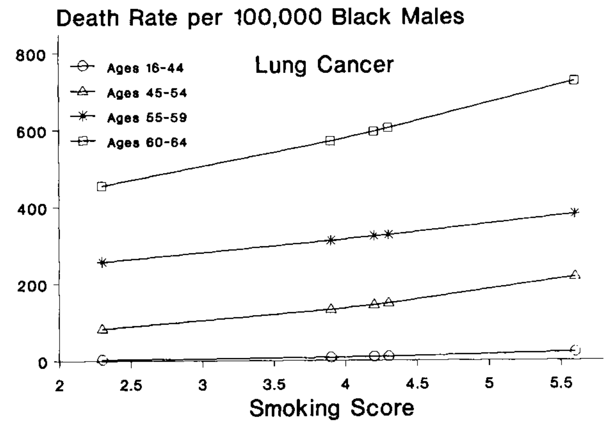

Fig. 5. Predicted death rates for lung cancer by smoking score and age in black males. Rates were annual rates per 100,000 estimated in the California Occupational Mortality Study. Alcohol and socioeconomic status were held constant at their median values. Source of smoking data for occupations: National Health Interview Survey, 1978–1980, Smoking Supplement Subsample.

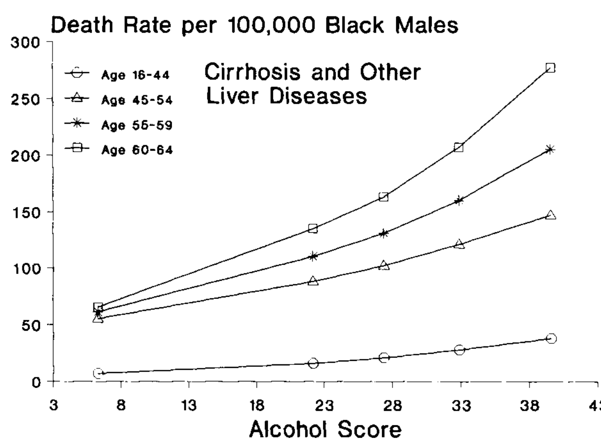

Fig. 6. Predicted death rates for cirrhosis of the liver by alcohol consumption score and age in black males. Rates were annual rates per 100,000 estimated in the California Occupational Mortality Study. Smoking and socioeconomic status were held constant at their median values. Source of alcohol consumption data for occupations: National Health Interview Survey, 1977, HI Supplement.

DISCUSSION

Adjustment for smoking, alcohol, and socioeconomic status caused the mortality ratios to be generally closer to 1.0, and judging from the effect on occupations where smoking and drinking habits are well-known, the adjustments appeared to work well. For example, the mortality ratio for bartenders (high smoking and drinking in the National Health Interview Survey data) for all causes of death combined changed from 2.0 before adjustment to 1.3 after adjustment. Similarly, but in the opposite direction, the mortality ratio for physicians and dentists (low smoking and drinking) for all causes combined changed from 0.4 before adjustment to 1.0 after adjustment. The effects of adjustment were usually moderate and similar to those of

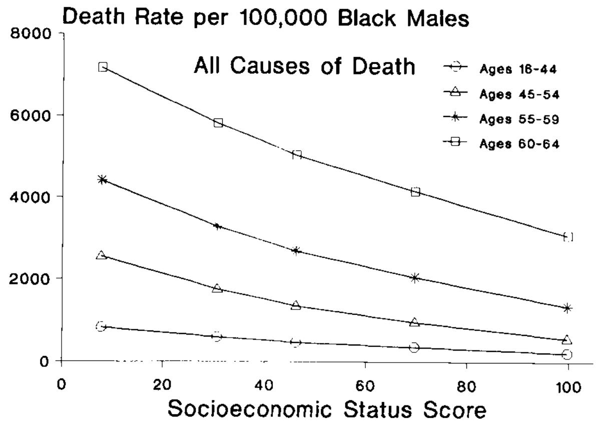

Fig. 7. Predicted death rates for all causes combined by socioeconomic status score and age in black males. Rates were annual rates per 100,000 estimated in the California Occupational Mortality Study. Smoking and alcohol were held constant at their median values. Source of socioeconomic status data for occupations: Nam-Powers Occupational Status Scores, Center for the Study of Population, Florida State University, 1986.

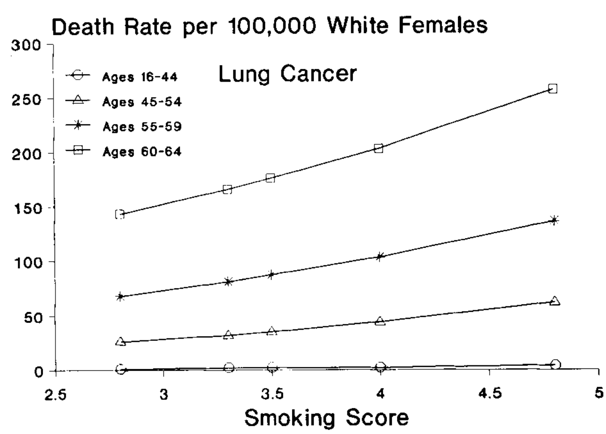

Fig. 8. Predicted death rates for lung cancer by smoking score and age in white females. Rates were annual rates per 100,000 estimated in the California Occupational Mortality Study. Alcohol and socioeconomic status were held constant at their median values. Source of smoking data for occupations: National Health Interview Survey, 1978–1980, Smoking Supplement Subsample.

studies that have compared unadjusted and adjusted risk ratios where data on individuals were available for smoking or socioeconomic status [Siemiatycki et al., 1988; Blair et al., 1985].

Limitations of the Methods and Data

The methods we used for adjustment had several limitations. One limitation was the time frame of the survey data; for example, the validity of using 1978–1980 smoking data relied on the assumption that occupational differentials in smoking prevalence have persisted over time. There is some evidence to support this assump-

Fig. 9. Predicted death rates for cirrhosis of the liver by alcohol consumption score and age in white females. Rates were annual rates per 100,000 estimated in the California Occupational Mortality Study. Smoking and socioeconomic status were held constant at their median values. Source of alcohol consumption data for occupations: National Health Interview Survey, 1977, H1 Supplement.

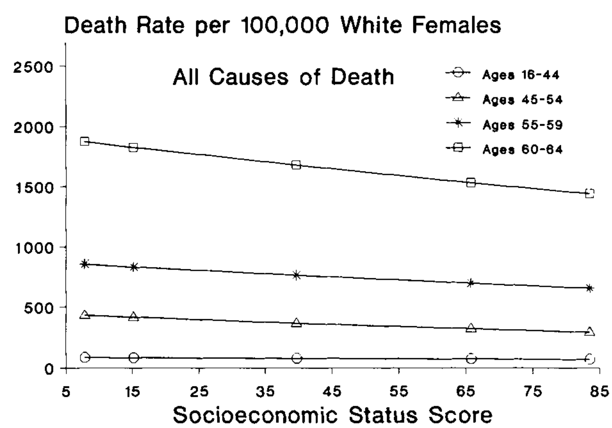

Fig. 10. Predicted death rates for all causes combined by socioeconomic status score and age in white females. Rates were annual rates per 100,000 estimated in the California Occupational Mortality Study. Smoking and alcohol were held constant at their median values. Source of socioeconmic status data for occupations: Nam-Powers Occupational Status Scores, Center for the Study of Population, Florida State University, 1986.

tion based on comparing broad occupational groups between 1970 and 1980 [Sterling and Weinkam, 1976]. However, the rate of decline in smoking prevalence has been faster among the more educated [Pierce et al., 1989]. The alcohol data shared these limitations, but the latent period is often short or nonexistent for deaths due to alcohol.

The scale of 1 to 10 we assigned to the smoking categories was somewhat arbitrary—it was based on the approximate relative risks for lung cancer—and may have been more appropriate for some causes of death than others. Alternatives that should be considered include a separate index for each cause of death, and an index based solely on quantity of cigarettes smoked, although "former smokers" are dif-

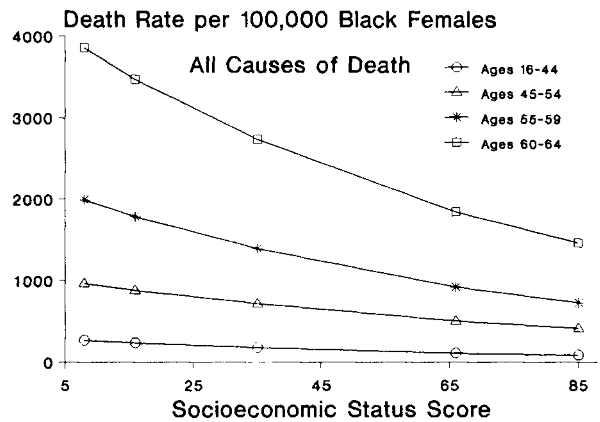

Fig. 11. Predicted death rates for all causes combined by socioeconomic status score and age in black females. Rates were annual rates per 100,000 estimated in the California Occupational Mortality Study. Source of socioeconomic status data for occupations: Nam-Powers Occupational Status Scores, Center for the Study of Population, Florida State University, 1986.

TABLE III. Tests for Significance of Models With Smoking, Alcohol Consumption, and Socioeconomic Status Compared to Baseline Models With Age Only*

		Test of significance of full model compared to baseline[a]					
		Lung cancer		Cirrhosis of liver		All deaths	
Population	Degrees of freedom	Chi^2	p	Chi^2	p	Chi^2	p
White male	12	633	0.000	1084	0.000	5277	0.000
Black male	12	138	0.000	274	0.000	1837	0.000
White female	12	285	0.000	122	0.000	577	0.000
Black female[b]	4	17	0.002	36	0.000	314	0.000

*Significance was assessed with the likelihood ratio Chi square test with degrees of freedom equal to the No. of additional parameters. Log likelihood statistics were calculated during Poisson regression of death rates in the California Occupational Mortality Study.
[a]Full models included age, smoking, alcohol, socioeconomic status, and parameters for 2-way interaction between age and the other variables (the four age groups were indicated by three categorical variables).
[b]Black female full models included only age and socioeconomic status due to small sample sizes for black female occupations in National Health Interview Surveys.

ficult to classify in such a scheme. The NHIS data that we used did not distinguish between heavy and light former smokers and did not consider length of time since quitting smoking. We also did not consider smoking differences by industry: smoking can vary by industry within occupation as well as by occupation within industry.

The Nam-Powers socioeconomic status scores were subject to several limitations: 1) they were based on individual rather than household education and income; 2) education and income were combined into a single index, when we would have preferred to adjust for each variable separately; and 3) the scores were not age-, sex-, or race-specific. The lack of specificity could have caused a small amount of bias in the occupational risk estimates [Rosenbaum and Rubin, 1984]. We chose Nam-Powers socioeconomic status scores over other existing occupational status scores because they have a simple interpretation (level of education and income) and do not

TABLE IV. Standardized Mortality Ratios for Agricultural Occupations Unadjusted and Adjusted for Smoking, Alcohol, and Socioeconomic Status, by Occupation and Cause of Death*

			Cause of death				Cancer																	
Occupation	Standardized mortality ratio		All causes	Hypertension	Ischemic heart disease	Other circulatory disease	Lip, oral	Stomach	Colon	Digestive organs	Trachea, bronchus, and lung	Genital organs	Urinary organs	Leukemia	Lymphoid tissue	Other sites	Cerebrovascular disease	Chronic obstructive pulmonary disease	Cirrhosis of liver	Urinary system	Falls and machinery accidents	Other accidents	Suicide	All other causes
Farmers		Hi[a]	110	119	122	104	199	119	138	132	92	268	182	153	200	149	100	161	89	327	322	207	129	137
	Unadjusted	SMR	104	63	108	84	101	51	85	92	72−	169	104	74	125	107	65	111	65−	166	208+	175+	92	115
		Low	97	29	95	66	43	16	49	62	55	100	54	30	73	75	40	73	45	71	127	147	63	96
		Hi	129	136	140	135	272	132	141	164	137	219	233	135	198	183	106	241	124	325	272	201	171	148
	Adjusted	SMR	121+	72	124+	108	138	57	87	115	108	138	133	65	124	132	69	165+	90	165	176+	171+	121	124+
		Low	114	33	109	85	59	18	49	77	83	82	69	26	72	93	42	109	63	71	108	144	84	104
	Deaths		969	9	257	77	8	5	16	29	65	18	12	7	17	36	20	27	36	8	20	140	33	129
Farm workers		Hi	173	129	112	112	199	176	81	133	101	187	89	194	81	126	228	195	253	360	468	332	98	277
	Unadjusted	SMR	166+	81	101	94	119	108	47−	100	84	119	43−	128	44−	96	185+	147+	221+	225+	380+	310+	79−	257+
		Low	160	47	90	79	65	62	25	73	69	70	17	80	21	72	148	108	193	131	305	289	63	237
		Hi	145	116	114	111	189	156	103	135	112	154	104	172	85	137	176	161	180	210	213	210	110	180
	Adjusted	SMR	140+	72	103	93	113	96	60	101	92	97	50	113	46−	105	143+	121	158+	131	173+	196+	88	167+
		Low	134	42	92	78	62	55	32	74	76	58	20	71	22	78	114	89	138	76	139	183	70	154
	Deaths		2805	17	347	132	14	16	13	45	107	18	7	22	10	52	87	47	220	17	89	821	82	642
Gardeners other agric. workers		Hi	107	156	72	110	146	223	127	119	92	167	149	152	218	118	131	113	172	319	373	138	160	159
	Unadjusted	SMR	101	87	60−	87	57	124	73	78	71−	85	72	77	140	81	88	66	138+	155	272+	119+	128+	139+
		Low	95	44	50	68	16	62	38	48	53	37	29	33	84	53	57	35	109	62	192	103	102	121
		Hi	88	133	72	98	118	188	181	111	90	163	170	155	245	112	100	83	107	191	238	101	149	103
	Adjusted	SMR	83−	74	61−	78−	46	105	104	72	69−	83	83	78	157	77	68	48−	86	93	173+	88	120	90
		Low	78	37	50	61	13	52	54	45	52	36	33	34	94	50	44	26	68	37	123	76	95	78
	Deaths		1021	11	124	73	4	11	12	21	54	8	7	8	19	26	25	13	80	7	38	191	80	209

*White males, California Occupational Mortality Study.

Adjusted expected death rates were calculated with prediction models based on Poisson regression of observed death rates for occupations on age and estimated levels of smoking, alcohol, and socioeconomic status. Standardized mortality ratios were calculated by dividing the observed no. of deaths by the expected no. and multiplying by 100.

−, the upper 95% confidence limit is less than 100 (the mortality ratio is significantly low).

+, the lower 95% confidence limit is greater than 100 (the mortality ratio is significantly high).

[a]Hi/Low indicates upper and lower 95% confidence limits.

involve subjective evaluations of socioeconomic status [Nam and Terrie, 1986; Liberatos et al., 1988].

A limitation common to all three adjustments (smoking, alcohol, and socioeconomic status) was projection of national estimates to Californians. While there is evidence that smoking prevalence varies between states [Marks et al., 1985], occupational differentials are likely to be similar in most large states.

A potential source of bias was estimation of confounder effects without occupation in the models. The bias would occur when an occupational exposure and a confounder were correlated. For example, if asbestos exposure and smoking were highly correlated, then adjusting for smoking would partially adjust for asbestos, and mortality ratios for high asbestos occupations would be biased towards 1.0. Of particular concern was correlation of socioeconomic status and hazardous exposure. We did not include occupation in the models when estimating confounder effects because of the difficulty of quantifying "occupation" as an exposure variable. For example, in our regression models for males, 67 dummy variables would have been required to represent the 68 occupations, and the total number of parameters (82 including parameters for potential confounders and interactions) would have been very large relative to the number of observations (268). The potential effect of not including occupation in our models would be that our SMRs are conservative, i.e., that they underestimate occupational risk.

Better data on potential confounders would improve occupational risk estimates. In particular, we recommend that sample sizes be made larger for nonwhites in surveys of lifestyle factors. We also recommend further research into the methods of controlling for lifestyle factors, as follows: 1) evaluation of other data on death certificates that might be predictive of mortality and be distributed differently between occupations, such as marital status or urban versus rural county of residence; 2) examination of the effects of socioeconomic status on mortality within categories of age, gender, and race; 3) further exploration of the relationship between smoking, drinking, education, and income; 4) application of the adjustment methods to proportionate mortality analyses which do not depend on census occupation estimates; and 5) application of the empirical Bayes procedure to the mortality ratios themselves so that mortality ratios with small numbers of deaths are adjusted toward the average.

Limitations of Death Certificate Data

Death certificate data have limitations that have been well described by other investigators [Melius et al., 1989; Dubrow et al., 1987]. Briefly, the major limitations relate to accuracy of the occupation and cause of death statements, differences between death certificate and census ascertainment of occupation, and lack of information on potential confounders such as smoking. Because the accuracy of the cause of death and occupation on certificates of death may be better for younger decedents, the California Occupational Mortality Study was limited to ages 16–64.

The adjusted mortality rate ratios should be interpreted cautiously. To assess their sensitivity to method of calculation, they should be interpreted in conjunction with rate ratios adjusted for age only (or age and time if a long period of time is considered) and in conjunction with proportionate mortality ratios that do not rely on census estimates of occupational populations. Because of their limitations the adjusted mortality ratios should be used primarily to develop hypotheses for further study of occupational and nonoccupational risk factors and not be interpreted as definitive evidence of cause and effect between occupation and disease or injury.

To our knowledge this is the first study to adjust occupational mortality surveillance data using National Health Interview Survey and U.S. Census data to describe the smoking, alcohol, and socioeconomic status of occupations. The adjustment procedures we chose were among many possible procedures, and we encourage further experimentation with use of these valuable data.

ACKNOWLEDGMENTS

We would like to express our appreciation for assistance and guidance from Linda Rudolph, M.D., California Department of Health Services; Thomas S. Scopp, Bureau of the Census; Charles Nam, Ph.D., Florida State University; and Steven Samuels, Ph.D., and LaVon Missell, M.S., Division of Occupational Medicine, University of California, Davis. We would also like to acknowledge helpful comments from Jane Norbeck, Neil Maizlish, Eva Glazer, Richard J. Jackson, Richard G. Ames, and Bruce B. Cohen.

This study was supported by California Department of Health Services Interagency Master Agreement No. 85-87171.

REFERENCES

Blair A, Hoar SK, Walrath J (1985): Comparison of crude and smoking-adjusted standardized mortality ratios. J Occup Med 27:881–884.

Brackbill R, Frazier T, Shilling S (1988): Smoking characteristics of U.S. workers, 1978–1980. Am J Ind Med 13:5–41.

California Department of Health Services (1987): "California Occupational Mortality, 1979–81." Sacramento, CA: CDHS Health Demographics Section, March.

Dubrow R, Sestito JP, Lalich NR, Burnett CA, Salg JA (1987): Death certificate-based occupational mortality surveillance in the United States. Am J Ind Med 11:329–342.

Efron B, Morris C (1975): Data analysis using Stein's estimator and its generalizations. J Am Stat Assoc 70:311–319.

Efron B, Morris C (1977): Stein's paradox in statistics. Sci Am 236:119–127.

Frome EL, Checkoway H (1985): Use of Poisson regression models in estimating incidence rates and ratios. Am J Epidemiol 121:309–323.

Liberatos P, Link BG, Kelsey JL (1988): The measurement of social class in epidemiology. Epidemiol Rev 10:87–120.

Lovett AA, Bentham CG, Flowerdew R (1986): Analysing geographic variations in mortality using Poisson regression: the example of ischemic heart disease in England and Wales 1969–1973. Soc Sci Med 23:935–943.

Manton KG, Stallard E, Woodbury MA, Riggan WB, Creason JP, Mason TJ (1987): Statistically adjusted estimates of geographic mortality profiles. JNCI 78:805–815.

Marks JS, Hogelin GC, Gentry EM, Jones JT, Gaines KL, Forman MR, and Trowbridge FL (1985): The behavioral risk factor surveys: I. state-specific prevalence estimates of behavioral risk factors. Am J Prev Med 1:1–8.

Melius JM, Sestito JP, Seligman PJ (1989): IX. Occupational disease surveillance with existing data sources. Am J Public Health [Suppl] 79:46–52.

Morris CN (1983): Parametric empirical Bayes inference: theory and applications. J Am Stat Assoc 78:47–55.

Nam CB, Powers MG (1983): "The Socioeconomic Approach to Status Measurement." Houston: Cap and Gown Press.

Nam CB, Terrie EW (1986): Comparing the Nam-Powers and Duncan SEI Occupational scores. Working paper, Center for the Study of Population, Florida State University. WPS 86-27, ISSN 0740-9095.

National Center for Health Statistics (1978): "Current Estimates from the Health Interview Survey: United States- 1977." Series 10 no. 126. DHEW Pub. no. (PHS) 78-1554.
Pierce JP, Fiore MC, Novotny TE, Hatziandreu EJ, Davis RM (1989): Trends in cigarette smoking in the United States: educational differences are increasing. JAMA 261:56–60.
Rosenbaum PR, Rubin DB (1984): Difficulties with regression analyses of age-adjusted rates. Biometrics 40:437–443.
Rothman KJ, Boice JD (1979): "Epidemiologic Analysis with a Programmable Calculator." NIH Publication No. 70-1649, 1979.
Schoenborn CA, Danchik KM (1980): Health practices among adults: United States, 1977. National Center for Health Statistics, Advance data, no. 64, HHS Pub. no. (PHS) 81-1250.
Siemiatycki J, Wacholder S, Dewar R, Cardis E, Greenwood C, Richardson L (1988): Degree of confounding bias related to smoking, ethnic group, and socioeconomic status in estimates of the associations between occupation and cancer. J Occup Med 30:617–625.
Singleton JA, Beaumont JJ (1989): COMS II—California Occupational Mortality, 1979–1981, Adjusted for Smoking, Alcohol, and Socioeconomic Status. Sacramento, CA: California Department of Health Services, December.
Singleton JA, Beaumont JJ, Doebbert GD (1989): A computer program for analyses of vital statistics-based occupational mortality data. Comput Biomed Res 22:488–496.
Sterling TD, Weinkam JJ (1976): Smoking characteristics by type of employment. J Occup Med 18: 743–754.
Williams RL, Cunningham GC, Norris FD, Tashiro M (1980): Monitoring perinatal mortality rates: California, 1970 to 1976. Am J Obstet Gynecol 136:559–568.

TECHNICAL APPENDIX

Calculation of Final Smoking and Alcohol Scores

For the empirical Bayes adjustment it was assumed that, conditional on the actual smoking or alcohol scores S_i, the observed preliminary scores s_i were independently and normally distributed with means S_i and variances V_i, where i denotes occupation. The V_i represented the variation associated with taking a sample to estimate the actual score. The actual scores were also considered as random variables with a normal distribution having mean U and variance A, where U was the actual score for all occupations combined and A was the variation of actual scores between occupations. It can then be shown that the distributions of the actual scores S_i conditional on observing the preliminary scores s_i are also normally distributed with means $E\{S_i| s_i\} = (1-B_i)*s_i + B_i*U$ and variances $V_i*(1-B_i)$, where the $B_i = V_i/(V_i + A)$ are the shrinkage parameters [Morris, 1983].

Hence the goal of estimating the $E\{S_i|s_i\}$ was accomplished by estimating B_i and U. The B_i were estimated by $b_i = ((N-3)/(N-1))*V_i/(V_i+A)$, where i runs from one to N (the number of occupations). The overall score U was calculated as the weighted average of the occupation-specific scores, where the weights were the estimated national populations represented by the survey respondents in each occupation. The unknown quantity A, the variation of actual scores between occupations, was estimated by

$$a = \frac{(\text{sum over i of } (((N/(N - 1))*(s_i - U)^2) - V_i)/(V_i + A))}{(\text{sum over i of } 1/(V_i + A))} . \tag{1}$$

Since the estimate of A included A in the sums on the right side of equation (1), an iterative solution was required. For a starting value of A we used

$$a = (\text{sum over i of } V_i{*}(s_i - U)^2)/(N-3){*}(\text{sum over i of } V_i)$$

and repeated formula 1 until successive estimates of A were the same up to seven decimal places. The final estimates of A and U were used to calculate the estimated shrinkage parameters b_i.

Occupations were left out of the empirical Bayes procedure when the number of people sampled (n_i) was less than five, and were assigned the score for all occupations combined in the appropriate subgroup. The iterative calculation of the final scores was done using SAS, with convergence reached in seven iterations or less for each of the four gender by race subgroups.

The variances V_i of the preliminary alcohol scores, which were weighted proportions, were estimated by $s_i(1-s_i)/n_i$, where n_i was the number of survey respondents in occupation i. Variances for the preliminary smoking scores were estimated assuming a multinomial distribution for the proportions of never-smokers, former smokers, current smokers averaging less than 20 cigarettes per day, and current smokers averaging 20 or more cigarettes per day. If an estimated proportion was zero or one, a value of .05 and .95, respectively, was used for purposes of estimating the variance.

The final values of the between occupation variation (A) obtained iteratively are shown below:

	Smoking A	Alcohol A
White male	0.44	0.004
Black male	0.80	0.016
White female	0.24	0.002
Black female	0.00	0.000.