Diesel Exhaust Exposure and Lung Cancer: Adjustment for the Effect of Smoking in a Retrospective Cohort Study

Emma K. Larkin,[1] Thomas J. Smith,[2] Leslie Stayner,[3] Bernard Rosner,[5] Frank E. Speizer,[5] and Eric Garshick[1,4,5]*

Background *The extent that cigarette smoking may confound the relationship between diesel exhaust exposure and lung cancer was assessed in a retrospective cohort study of 55,395 U.S. railroad workers followed from 1959 to 1976.*

Methods *The relative risk (RR) of lung cancer due to diesel exhaust was indirectly adjusted using job-specific smoking data from a case-control study of railroad workers who died between 1981–1982 and from a survey of 514 living workers from an active railroad in 1982. Adjustment factors were developed based on the distribution of job-specific smoking rates.*

Results *The unadjusted RR for lung cancer was 1.58 (95% CI = 1.14–2.20) for workers aged 40–44 in 1959, who experienced the longest possible duration of exposure, and the smoking adjusted RR was 1.44 (1.01–2.05).*

Conclusions *After considering differences in smoking rates between workers exposed and unexposed to diesel exhaust in a relatively large blue-collar cohort, there were still elevated risks of lung cancer in workers in jobs with diesel exhaust exposure.* Am. J. Ind. Med. 38:399–409, 2000. *Published 2000 Wiley-Liss, Inc.*[†]

KEY WORDS: confounding; diesel exhaust; lung cancer; railroad workers; smoking

INTRODUCTION

A limitation of retrospective studies of lung cancer is the inability to directly obtain information regarding cigarette smoking. The degree that cigarette smoking confounds the results of a retrospective study depends on the extent to which the distribution of smoking behavior differs among subjects with and without the exposure under study. This limitation may be important if the true relative risk of disease is small, since differences in cigarette smoking may explain a substantial part of the observed relative risk attributable to the exposure. One method of controlling for smoking in retrospective studies is to compare the risk of disease among individuals of similar socioeconomic classes, i.e., blue collar workers, since smoking behavior is a correlate of socioeconomic class [Brackbill et al., 1988; Stellman et al., 1988]. In 1988, our research group reported the results of a retrospective cohort study conducted among 55,407 railroad workers that indicated that workers with the most diesel exhaust exposure had the greatest risk of dying of lung cancer [Garshick et al., 1988]. An internal control population of unexposed clerks and signal workers was selected to match the social class gradient of the exposed brakemen, conductors, hostlers, engineers, firemen, and shop workers.

It was not possible to assess the smoking habits of the cohort directly because mortality was ascertained retrospectively between 1959 and 1980.

We apply the distribution of job-specific smoking habits of occupations within the railroad industry to assess whether using an internal referent population effectively controls for the impact of smoking on lung cancer. These smoking histories were obtained from next-of-kin in an accompanying case-control study of deaths in the U.S. railroad industry collected between 3/1/81 and 2/28/82 [Garshick et al., 1987a], and smoking habits obtained from a survey of active railroad workers in 1982 [Garshick et al., 1987b]. We use the method of adjustment proposed by Schlesselman [1978] and Axelson [1980] who illustrated their methods by using hypothetical values for the proportion of smokers, whereas we use actual data from railroad workers and derive confidence limits for our adjusted estimates.

MATERIALS AND METHODS

Cohort Study

Details regarding the design of the retrospective cohort study have been previously published [Garshick et al., 1988]. Briefly, the U.S. railroad industry changed from steam to diesel power over the decade of the 1950s such that by 1952, roughly half the locomotives were diesel powered and by 1959, 95% were powered by diesel. The U.S. Railroad Retirement Board (RRB) maintains work, retirement, and death records for each railroad worker with more than 10 years of railroad employment. An Interstate Commerce Commission (ICC) job code was available yearly for each selected railroad worker starting in 1959 until death or retirement. Based on an industrial hygiene survey [Woskie et al., 1988a, 1988b], of over 150 job categories, 39 job codes were selected for study and classified as representing workers with either diesel-exposed (hostlers, firemen, brakemen, conductors, engineers or shop workers) or unexposed (signal maintainers or clerks) jobs. A cohort of 55,407 subjects aged 40–64 in 1959 with 10–20 years of railroad work was defined in 1959. Deaths were ascertained from RRB records through 1980, and death certificates were received for 88% of the known 19,396 deaths. Cases of lung cancer were identified using death certificates.

After publishing the initial results, it was recognized that there was an under-ascertainment of deaths between 1977 and 1980, with up to 70% of the deaths missing by 1980 [Crump, 1999; Crump et al., 1991; Health Effects Institute, 1995]. Therefore, in contrast to the original publication, the current analysis was limited to the years 1959–1976, where death ascertainment was believed to be complete, and 12 subjects were excluded due to erroneous work histories.

Poisson regression was performed using the statistical package EPICURE [Preston and Pierce, 1993]. Railroad workers tended to remain in the same job category throughout their careers; and, therefore, job title in 1959 was a predictor of future exposure category. Age in 1959 was divided into 5-year age groups: 45–49, 50–54, 55–59, 60–64, with age group 40–44 used as the referent category. Occupational codes were classified as diesel-exposed/unexposed in 1959 or stratified into the diesel exposure categories of engineers/fireman, brakemen/conductors/hostlers, and diesel repair shop workers. To assess the effect of working in a diesel-exposed job within each 1959 age group, interaction terms between age category in 1959 and job category were constructed. Thus, the regression model included the interaction terms between 5-year age groups in 1959 and 1959 job category as well as indicator variables for age group and job category. This allowed the calculation of the relative risk of workers in a diesel exhaust exposed job in 1959 relative to unexposed workers in the same age category. Calendar year in 1-year intervals and attained age in 5-year intervals were controlled for in the analysis using the stratification command in EPICURE. A log-linear relationship between the predictor variables and lung cancer mortality was assumed. In contrast to the results in the original publication, in this analysis, occupational code-specific relative risks (engineers/fireman, brakemen/conductors/hostlers, and diesel repair shop workers) are presented.

Additional analyses in this cohort have been conducted using other measures of cumulative exposure [California Environmental Protection Agency, 1998; Crump, 1999; Crump et al., 1991], such as years in a diesel exhaust-exposed job and levels of diesel particulate estimates based on the results on industrial hygiene surveys [Woskie et al., 1988a, 1988b]. However, these measures have limitations that would lead to exposure misclassification. It was not possible to determine the earliest that each subject had worked with diesel locomotives rather than steam locomotives before 1959. Due to differences in engine design, it is possible that work with diesel locomotives early in the follow-up period would have resulted in more diesel exposure than in later years. Finally, estimates of exposure to diesel were obtained from four smaller Northern railroads and may not be applicable to all exposure scenarios. Consequently in this paper, exposure will only be considered based on job title in 1959 as a surrogate of cumulative exposure.

Overview of Smoking Adjustment Method

The data for the adjustment procedure was obtained from two sources: next-of-kin smoking histories obtained from a case-control study of deaths among railroad workers who died between March 1981 and February 1982

[Garshick et al., 1987a] and a mail survey of 514 workers in a single railroad conducted in 1982 [Garshick et al., 1987b].

Case-Control Study

The original case-control study was designed as a matched case-control study of lung cancer and diesel exhaust exposure. Lung cancer deaths, in railroad workers born after 1899, were matched on age and date of birth with up to two randomly selected control deaths from the same cohort who died within 30 days of the case, after excluding those who died of an accidental cause or cancer. Additional deaths were selected from cases of other cancer and respiratory deaths.

Smoking data were collected through a questionnaire sent through the mail or a telephone interview with the next-of-kin of the decedent. Questions about smoking included: (1) Did the deceased ever smoke cigarettes? (2) How old was the deceased when he last smoked? (3) On the average of the entire time he smoked, how many cigarettes per day did he smoke? (4) About how old was he when he first started to smoke cigarettes? Smoking history was categorized into the following: (1) never smokers, (2) smokers of 1–20 cigarettes per day, (3) smokers of 21 or more cigarettes per day, (4) former smokers who had quit for at least 2 years but less than ten, and (5) former smokers who had quit for 10 or more years. Of proxy respondents 78.9% were spouses, and 11.9% were children of the deceased. The remaining respondents consisted of other relatives or family friends. Of 3,554 white male subjects with complete smoking histories and a known occupation in 1959, there were 883 individuals who died from lung cancer, 414 individuals who died of respiratory deaths, 717 individuals who died from other cancers, and 1,540 individuals who died of other causes except accidental deaths.

The smoking categories in the case-control study were stratified by the subject's age in 1959 to consider the smoking habits of workers in the same age as those in the retrospective cohort study. Job groups (engineer/fireman, conductor/brakeman/hostler, shop workers, and clerk and signal workers) were defined using the 39 job codes listed in the original cohort study [Garshick et al., 1987a, 1988]. Since the case-control study also included subjects with occupational codes in these job groups not selected for the cohort study [Garshick et al., 1987a], additional analyses were conducted with these other railroad job codes included in order to increase the number of subjects in each smoking category.

Mail Survey

Smoking habits in living railroad workers were based on 514 white males on the payroll of an active U.S. railroad who had responded to a mail survey in 1982 [Garshick et al., 1987b]. Smoking status was based on self-reported answers to the following questions [Ferris, 1978]: ''did you ever smoke,'' and ''do you smoke now?'' Ex-smokers were workers who had quit smoking more than 1 month before completing the questionnaire. Additional data were gathered on average cigarettes per day consumed and duration of smoking, including age at initiation and age upon cessation. When only ICC job codes used in the cohort study were used to define smoking categories, the sample was reduced to 220 white males. When all occupational codes included in the engineer/fireman, conductor/brakeman/hostler, shop workers, and clerk and signal worker job groups were used, the sample included 413 white males. Because sample sizes precluded analyses using more detailed definitions of smoking, smoking was considered in three categories: current smokers, former smokers, and never smokers.

Smoking Adjustment

A method proposed by Schlesselman [1978] and Axelson [1980] was used to adjust the relative risk of lung cancer for the effects of smoking. Smoking adjustment factors were obtained based on deceased (case-control) and living (mail survey) railroad workers. For each diesel exhaust exposure category (job group) a smoking-weighted relative risk was calculated by using the distribution of workers in each smoking category to weight literature-based relative risks [Burns et al., 1997; Hrubek et al., 1997] of lung cancer due to cigarette smoking, described in Appendix A. A smoking adjustment factor was calculated for each job group by dividing the smoking-weighted relative risk for workers exposed to diesel exhaust by the smoking-weighted relative risk calculated for unexposed workers. The adjusted relative risk for each job group was obtained by dividing the observed relative risk by the smoking adjustment factor (Appendix B). A method to derive confidence limits for this estimate of relative risk is also described in Appendix B. This method takes into account the sampling error in determining the smoking proportions among occupational groups. SAS version 6.12 [SAS Institute, 1996] was used to generate frequency distributions of cigarette smoking, and Microsoft Excel [Microsoft Corp., 1997] spreadsheets were used to apply the smoking adjustment methodology. Mean years of smoking, mean years since quitting, and cigarettes smoked per day were compared among job groups using either ANOVA or a Kruskal Wallis nonparametric test as appropriate.

RESULTS

Of the 55,395 railroad workers selected based on age and job category in 1959, 36% were in the youngest age stratum (40–44) and 8% were in the age stratum 60–64

TABLE I. Number of Workers and Lung Cancer Deaths in Each Occupational Group in a Cohort of 55,395 Railroad Workers with Lung Cancer Mortality Ascertained 1959–1976

| | Age in 1959 | | | | | |
	40–44	45–49	50–54	55–59	60–64	Total
Total	20,076 (36.2)	13,941 (25.2)	9,846 (17.8)	7,237 (13.1)	4,295 (7.8)	55,395
Unexposed						
Clerk/signal maintainers	4,966 (24.7)	3,139 (22.5)	2,603 (26.4)	1,994 (27.6)	1,321 (30.8)	14,023
Exposed						
Engineers/firemen	4,052 (20.2)	2,851 (20.5)	1,863 (18.9)	1,397 (19.3)	840 (19.6)	11,003
Brakemen/conductors/hostlers	7,352 (36.6)	5,109 (36.6)	2,911 (29.6)	1,923 (26.6)	982 (22.9)	18,277
Shop workers	3,706 (18.5)	2,842 (20.4)	2,469 (25.1)	1,923 (26.6)	1,152 (26.8)	12,092
Lung cancer deaths	248	295	314	338	194	1,389
Median retirement year[a]	1977	1974	1971	1967	1963	

[a]Retirement year based on follow-up through 1980.

(Table I). Approximately 69–77% of workers in each age stratum worked in jobs with potential diesel exhaust exposure since dieselization of the railroads did not occur until the mid-1950s. Younger workers would have had the opportunity for a greater duration of exposure, as evidenced by their later retirement year compared to the other groups. The older workers would have had the least opportunity to work in a diesel exhaust-exposed job after 1959. Based on work in a diesel exposed job in 1959 and within each of the three occupational groups classified as diesel exhaust exposed, workers age 40–44 in 1959 had the highest relative risk of dying of lung cancer, while the older workers had lower relative risks [Table IIa]. These results are similar to the findings of the original analysis conducted through 1980.

Since the case-control study excluded workers born before 1900, it was not possible to determine smoking frequencies for the workers age 60–64 in 1959. Based on job title in 1959, workers categorized as shop workers tended to smoke the same or less than the unexposed workers (clerks and signal maintainers, Table III), using the job codes in the cohort study. Engineers/firemen and brakemen/conductors/hostlers tended to have a greater prevalence of smoking compared to unexposed workers. There were more never-smokers among the older workers, consistent with increased early mortality in younger workers who smoke.

Table III also shows the smoking-weighted relative risk for lung cancer in the occupational groups and the smoking adjustment factor (ratio of exposed/unexposed) for each category of diesel exhaust exposure. Adjustment factors greater than 1 reflect the positive confounding effects of smoking and adjustment factors less than 1 reflect negative confounding, in which the workers in the exposed job categories smoke less than workers in the unexposed categories. The smoking adjustment factors for the engineers/firemen and brakemen/conductors/hostlers in the youngest age stratum are all above 1, and are greatest in the 50–54 age group. Shop workers generally have smoking adjustment factors equal to or less than 1. Among workers aged 40–44, additional analyses using all subjects in the occupational groups increased the clerk/signal group to 206 subjects (12.6% never smokers), and workers in the exposed group to 342 subjects (7.7% never smokers). Among workers aged 45–49, additional analyses using all subjects in the occupational groups increased the clerk/signal group to 298 subjects (12.4% never smokers), and workers in the exposed group to 342 subjects (12.2% never smokers). The prevalence of smoking and smoking adjustment factors similar to those presented in Table III were obtained when all job codes used in the case-control study were included (data not shown).

Within the smoking categories, mean years of smoking were determined for all current smokers, and stratified by ≤ 20 and ≥ 21 cigarettes per day and compared among the diesel exhaust exposure categories (clerk/signal, engineers/firemen, brakemen/conductors/ hostlers, and shop workers). A similar analysis was done for years quit among ex-smokers, and average cigarettes per day were compared within smokers and ex-smokers. Among subjects corresponding to workers in the cohort aged 40–44 and 45–49, the workers with the greatest risk of lung cancer, there were small differences in smoking behavior (cigarettes per day and years smoked) that were not significant ($P = 0.08$-0.80) across occupational groups. However, within the subgroup of subjects who had quit smoking ≥ 10 years and using job

TABLE IIa. Relative Risk (RR) of Lung Cancer due to Diesel Exhaust Exposure by Occupational Group in a Cohort of 55,395 Railroad Workers Followed from 1959–1976

| | Occupational groups[a] | | | | | | | |
| | Engineers and firemen | | Brakemen, conductors, hostlers | | Shop workers | | Total exposed | |
Age in 1959 (years)	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI
40–44	1.83*	1.24–2.69	1.59*	1.11–2.27	1.32	0.86–2.01	1.58*	1.14–2.20
45–49	1.31	0.91–1.87	1.28	0.93–1.77	1.23	0.86–1.76	1.28	0.95–1.71
50–54	1.65*	1.20–2.26	1.15	0.84–1.56	0.97	0.69–1.35	1.21	0.93–1.58
55–59	1.01	0.72–1.41	1.24	0.92–1.65	1.12	0.83–1.50	1.13	0.87–1.43
60–64	0.72	0.46–1.13	1.30	0.90–1.87	0.93	0.64–1.35	0.98	0.72–1.34

[a]Results are from Poisson regression controlling for attained age and calendar year.
*$P < 0.05$.

TABLE IIb. Relative Risk (RR) of Lung Cancer due to Diesel Exhaust Exposure in a Cohort of 55,395 Railroad Workers Followed from 1959 to 1976 After Smoking Adjustment has been Applied

| | Occupational groups[a] | | | | | | | |
| | Engineers and firemen | | Brakemen, conductors, hostlers | | Shop workers | | Total exposed | |
Age in 1959 (years)	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI
40–44	1.61*	1.07–2.44	1.45	0.99–2.14	1.27	0.79–2.04	1.44*	1.01–2.05
45–49	1.17	0.79–1.74	1.08	0.76–1.54	1.21	0.80–1.83	1.12	0.81–1.54
50–54	1.30	0.91–1.86	0.95	0.67–1.36	1.05	0.70–1.57	1.04	0.77–1.41
55–59	1.05	0.71–1.56	1.24	0.88–1.76	1.29	0.86–1.94	1.18	0.92–1.51

[a]Original Poisson regressions adjusted for calendar year and attained age.
*$P < 0.05$.

codes selected for inclusion in the cohort study, differences were statistically significant ($P = 0.01$); unexposed subjects had smoked a mean of 19.9 years (n = 18), compared to subjects in the engineer/fireman group (29.0 years, n = 21), the brakeman/conductor group (27.9 years, n = 40), and the shop workers (25.1 years, n = 8). When all job codes were considered in subjects who had quit smoking ≥ 10 years ago, mean years of smoking were similar (clerk/signal 25.6 years, n = 49; engineers/firemen 29.0 years, n = 21; brakemen/conductors/hostlers, 27.9 years, n = 41; shop workers 23.7 years n = 18; $P = 0.26$).

The Poisson regression estimates for relative risk of diesel exhaust and lung cancer were divided by the adjustment factor to yield the new adjusted relative risk. Adjusting for cigarette smoking reduced the estimates of relative risk for workers aged 40–44 in each occupational group by 4–12% (Table IIb). The estimate for the engineers/firemen in the age group 50–54 in 1959 fell from 1.65 to 1.30 (21%). The relative risk of dying of lung cancer in workers aged 40–44 in the job group engineers/firemen and brakemen/conductors remained elevated, as did the overall risk for this age group after adjustment for cigarette smoking.

Table IV shows the smoking characteristics of railroad employees who were surveyed in 1982. Because of sample size constraints, it was only possible to stratify age into two groups: ≤ 50 years old and > 50 years old, and smoking into three groups: never, current, and ex-smokers. For similar reasons, engineers/firemen and brakemen/conductors/hostlers were combined into one exposure category. The ICC codes used were the job codes used in the retrospective cohort study for all occupational groups. For workers > 50, the adjustment factor for the exposed workers was less than 1, whereas for the workers ≤ 50,

TABLE III. Percent of Individuals in Each Smoking Category by Occupation and the Weighted Relative Risk of Smoking on Lung Cancer Mortality in a Sample of Railroad Workers who Died Between March 1981 and February 1982

| | Age in 1959 | | | | | | | | | |
| | 40−44 | | | | | 45−49 | | | | |
Sample characteristics[a]	Clerk, signal	Engineers, firemen	Brakemen, conductors, hostlers	Shop workers	Total exposed[b]	Clerk, signal	Engineers, firemen	Brakemen, conductors, hostler	Shop workers	Total exposed[b]
Cigarettes per day:										
None	14.5%	5.8%	7.7%	9.4%	7.3%	16.3%	12.8%	10.0%	10.5%	11.1%
20 or fewer	15.8%	16.5%	21.4%	25.0%	20.1%	26.7%	23.1%	24.4%	26.3%	24.1%
21+	27.6%	33.0%	31.0%	25.0%	31.0%	17.4%	23.1%	26.9%	10.5%	23.5%
Ex-smoker quit < 10 yrs	18.4%	24.3%	15.5%	15.6%	18.5%	7.0%	16.2%	15.6%	29.0%	17.5%
Ex-smoker quit ≥ 10 yrs	23.7%	20.4%	24.4%	25.0%	23.1%	32.6%	24.8%	23.1%	23.7%	23.8%
Sample size	76	103	168	32	303	86	117	160	38	315
Smoking weighted RR[c]	11.24	12.74	12.25	11.64	12.35	11.22	12.50	13.37	11.40	12.81
Adjustment factor[d]	1.000	1.133	1.090	1.035	1.099	1.000	1.114	1.192	1.017	1.142
	Age in 1959									
	50−54					55−59				
Cigarettes per day:										
None	19.8%	8.4%	13.0%	28.6%	15.0%	18.0%	14.0%	16.7%	35.6%	19.9%
20 or fewer	15.8%	26.5%	23.2%	16.1%	22.7%	21.4%	17.2%	22.5%	17.0%	19.5%
21+	13.9%	20.5%	21.3%	12.5%	19.0%	11.2%	9.7%	8.3%	10.2%	9.2%
Ex-smoker quit < 10 yrs.	17.8%	19.3%	13.9%	17.9%	16.6%	10.1%	15.1%	15.0%	10.2%	14.0%
Ex-smoker quit ≥ 10 yrs.	32.7%	25.3%	28.7%	25.0%	26.7%	39.3%	44.1%	37.5%	27.1%	37.5%
Sample size	101	83	108	56	247	89	93	120	59	272
Smoking weighted RR[c]	10.06	12.74	12.09	9.34	11.68	8.47	8.18	8.45	7.34	8.12
Adjustment factor[d]	1.000	1.267	1.202	0.929	1.162	1.000	0.966	0.997	0.866	0.958

[a]Occupational groups include only ICC job codes that are used in the cohort study.
[b]Total exposed consists of engineers, firemen, brakemen, conductors, hostlers, and shop workers combined.
[c]Weighted RR is the relative risk of lung cancer due to smoking weighted by the proportion of smokers in each smoking category. See Appendix A for the actual relative risks used in the calculation.
[d]Adjustment factor is the ratio of the weighted relative risk for diesel-exposed and diesel-unexposed workers, representing the extent to which different smoking habits may confound the relationship between diesel exhaust exposure and lung cancer.

the adjustment factor for the engineers/firemen and brakemen/conductors/hostlers was 1.138, values similar to those reported in Table III. There were no significant differences in cigarettes smoked per day, years of smoking, or in years since quitting smoking among occupational groups within each age group ($P = 0.08 - 0.79$). If these smoking adjustment factors were applied to the relative risks in Table IIa, the estimates of relative risk for lung cancer due to diesel exhaust exposure would be comparable to the relative risks presented in Table IIb. Furthermore, adjustment factors

similar to those presented in Table IV were obtained when all job codes in the case-control study were included (data not shown).

DISCUSSION

The relative risk of lung cancer mortality due to work in a job with diesel exhaust exposure in a retrospective cohort study in U.S. railroad workers was adjusted for the effects of cigarette smoking using job-specific smoking information

TABLE IV. Percent of Individuals in Each Smoking Category and the Weighted Relative Risk (RR) for Lung Cancer due to Smoking in a Population of Living Railroad Workers from an Active Railroad in 1982

Sample characteristics[a]	Current age (years)							
	≤ 50				> 50			
	Clerk, signal	Engineers, conductors[b]	Shop workers	Total exposed[c]	Clerk, Signal	Engineers, Conductors[b]	Shop workers	Total exposed[c]
Never	23.5%	13.5%	41.4%	21.4%	12.2%	15.9%	33.3%	20.3%
Current	35.3%	43.2%	20.7%	36.9%	34.2%	31.8%	26.7%	30.5%
Ex-Smoker	41.2%	43.2%	37.9%	41.8%	53.7%	52.3%	40.0%	49.2%
Sample size	17	74	29	103	41	44	15	59
Smoking weighted RR[d]	11.33	12.89	8.50	11.66	10.36	9.85	8.20	9.43
Adjustment factor[e]	1.000	1.138	0.750	1.03	1.000	0.951	0.791	0.909

[a]All occupational groups consist of ICC job codes that are used in the cohort study.
[b]Engineers and conductors also include firemen, brakemen and hostlers.
[c]Total exposed consists of engineers, firemen, brakemen, conductors, hostlers, and shop workers combined.
[d]Weighted RR is the relative risk of lung cancer due to smoking weighted by the proportion of smokers in each smoking category. See Appendix A for the actual relative risks used in the calculation.
[e]Adjustment factor is the ratio of the weighted relative risk for diesel-exposed and diesel-unexposed workers, representing the extent to which different smoking habits may confound the relationship between diesel exhaust exposure and lung cancer.

available from other railroad workers. Among younger workers (ages 40–44 in 1959) who had died in 1981–1982, engineers/firemen and brakemen/conductors/hostlers in 1959 had a slightly greater prevalence of smoking compared to workers unexposed to diesel exhaust. When adjustment factors that incorporated the differences in smoking prevalences were applied to the relative risk of lung cancer attributable to work in a diesel exhaust exposed job, the risk remained elevated among the younger workers in the cohort study. Smoking adjustment factors obtained from live railroad workers were similar to smoking adjustment factors obtained from deceased workers.

A potential weakness of this methodology applied to the retrospective cohort study relates to the sources of information on smoking. The effect of changing cigarette habits over time and its impact on mortality cannot be considered directly in the retrospective cohort study. Using deceased workers also overestimates the proportion of smokers in both the exposed and unexposed categories, since cigarette smoking is a risk factor for numerous causes of death. However, it is unlikely that the smoking behavior between 1959 and 1976 changed differently among the exposure groups. With the exception of one group, the number of cigarettes smoked and years since quitting smoking were similar among the exposure groups suggesting that smoking behavior was similar and unlikely to account for the overall effects of working in a diesel exhaust-exposed job. The adjustment factor would also not be altered due to the inclusion of the smoking histories

of deceased workers because the factor comprises the ratio of smoking behaviors in exposed and unexposed categories.

The effects of using a deceased population to estimate the effect of smoking in the cohort study also depends on whether there is an interaction between smoking and diesel exhaust on lung cancer. The proportion of deceased smokers in the diesel exhaust-exposed jobs would be greater if interactions were indeed present and, therefore, would exaggerate the true differences between exposure-specific smoking rates among living workers. Implicit in using the same relative risks for lung cancer mortality due to smoking for the diesel exhaust-exposed and unexposed categories is the assumption that there is no interaction between diesel exhaust and smoking that would modify the risk of lung cancer associated with smoking. The indirect adjustment methodology would then be inappropriate because we would not have accurate relative risks for the joint effect of smoking and diesel exposure on the risk of lung cancer mortality for use in the calculations. Emmelin et al. [1993] present limited data suggestive of a more than additive interaction, but their study lacked the power to be conclusive. Overall, there is insufficient information in the literature to assess whether there is a more than additive interaction between diesel exhaust and smoking in the occurrence of lung cancer.

Information on cigarette smoking from the case-control study was available from surrogate responders rather than the individuals whose deaths were reported. However,

Rogot and Reid [1975] found that surrogates were able to provide complete agreement in 92% of cases, differentiating between regular smokers and occasional or non-smokers. There was a tendency for the surrogate to overestimate the amount smoked, and agreement on amount smoked was present in 74%. Kolonel et al. [1977] found that a surrogate was in agreement 96% of the time regarding ever smoking, 84% of the time within 4 years of starting smoking, and 78% were within 10 cigarettes per day in estimating cigarette consumption. The mean number of cigarettes smoked per day reported by the surrogate was also greater than that reported by subject. Lerchen and Sanet [1986] noted perfect agreement among husbands and wives in reporting ever vs. never smoking cigarettes. There was no difference in reporting means years of smoking, age started smoking, and average number of cigarettes reported per day. McLaughlin et al. [1987] reported the accuracy of smoking histories first reported in 1971–1975 in men, and recalled by spouses and the initial respondent in 1982–1984. Spouses were able to report whether a person was an ever-smoker or non-smoker in 92% of cases, which was similar to the recall of the individual himself. Mean years of smoking were similar as well. Of 10,226 household proxy–respondent pairs, current smokers were identified correctly 96% of the time; ex-smokers (collapsing years-quit) were identified correctly 87% of the time, and never-smokers 81% of the time [Hyland et al., 1997]. Therefore, proxy respondents are generally able to report cigarette smoking status, and tend not to underestimate consumption. The extent to which smoking status is misclassified is unlikely to differ based on diesel exhaust exposure category.

Only limited information on smoking in railroad workers is available from the literature. Brackbill et al. [1988] from the National Health Information Survey (NHIS) reported that in currently employed railroad workers 31.5% never smoked, 44.3% currently smoked and 24.2% were former smokers in 1978–1980, but they did not distinguish among railroad jobs in detail. However, in 703 rail conductors included in the survey 40.7% were current smokers and 38.4% were never smokers. Stellman et al. used the American Cancer Society Prevention Study II to characterize the smoking habits across different occupational groups in 1982 [Stellman et al., 1988]. In a sample of 1,166 railroad workers they found that 19.7% were never smokers, 33.6% were current cigarette, pipe, or cigar smokers, and 47% were former smokers. The greater proportion of never smokers in these studies compared to our case-control study is consistent with the increased mortality attributable to smoking. The data from individuals who died between 1981 and 1982 closely represent the smoking habits of workers who died in the actual retrospective cohort between 1959 and 1976.

It is possible that risk factors for lung cancer other than smoking are associated with job category. Asbestos was used in the engine repair shops and the shop workers had the potential for exposure [Garshick et al., 1987b]. Although job categories with the potential for the heaviest shop asbestos exposure such as boiler makers were excluded from the cohort [Garshick et al., 1988], it is likely that other shop workers included in the cohort were exposed to asbestos, which could contribute to the risk of lung cancer in this group. Dietary factors [Albanes et al., 1996; Fontham, 1997; Omenn et al., 1996; Veierod et al., 1997] and other factors associated with lifestyle [Bandera et al., 1992; Knekt et al., 1991] have also been associated with lung cancer. However, it is unlikely that there are such significant dietary and lifestyle differences among subjects in this blue-collar cohort that would account for an elevated risk of lung cancer. We conclude that the original design of the cohort study was sufficient to eliminate the major effects of cigarette smoking as a confounder through the use of clerks and signal workers as an internal referent population. This conclusion is consistent with Siemiatycki et al. [1988] who also found that differences in smoking prevalence for lung cancer among various blue collar populations and a reference general population is not likely to have an adjustment factor of more than 1.2, and at most 1.3. Correction factors obtained in this study were generally less than 1.2. A study by Blair et al. [1988] examined the effects of smoking on the standardized mortality ratio for lung cancer, based on differences in smoking proportions among conductors and locomotive engineers compared to the general population. They report a small increase in the standardized mortality ratio from 123 to 128, further indicating that the absence of smoking data does not seriously confound this study. While the relative risk of lung cancer associated with diesel exhaust exposure is small, the effect of residual confounding is possible, though unlikely. Comparison of the effects of diesel exhaust exposure within a blue-collar population as done in the cohort study remains an efficient way to minimize the potential confounding effects of smoking.

As presented in this paper, the indirect adjustment of the data that consider the effects of smoking adds confidence to the association found between diesel exhaust exposure and lung cancer. An overall limitation of the literature describing the relationship between lung cancer and diesel exposure is a lack of a cohort with a long duration of exposure and follow-up (>20–30 years), together with a detailed exposure assessment that specifically measured diesel exhaust. Since occupational lung cancer typically develops over many years, the finding of a relationship between lung cancer and diesel exhaust exposure in such a cohort would lend considerable weight in support of declaring diesel exhaust a definite human lung carcinogen.

APPENDIX A: Relative Risks for Lung Cancer due to Smoking in White Males used to Determine its Potential Confounding Effect on the Association Between Diesel Exhaust and Lung Cancer

Smoking status	Age at death[a]			
	62−66	67−71	72−76	77−81
Deceased Railroad Workers (Case-Control Study):				
Non-smoker	1	1	1	1
1−20 cigarettes per day	12.9	14.8	14.8	14.8
21+cigarettes per day	19.3	23.4	23.4	23.4
Ex-smoker <10 yrs quit	11.5	11.5	11.5	11.5
Ex-smoker ≥ 10 yrs quit	6.8	6.8	6.8	6.8

[a]Each age range selected by ageing the cohort divided by age in 1959 (40−44, 45−49, 50−54, 55−59) by 22 years.

Smoking status	Age	
	≤ 50	>50
Live Railroad Workers (Mail Survey):		
Non-smoker	1	1
Current smoker	18	19.3
Ex-smoker	11.5	6.8

1. Data for current smokers taken from: Burns D, Shanks T, Choi W, Thun M, Heath C, Garfinkel L. The American Cancer Society cancer prevention study I: 12-year followup of 1 million men and women. In: Burns D, Garfinkel L, Samet J, eds., Changes in cigarette-related disease risks and their implication for prevention and control. National Institutes of Health, 1997: p157.

2. Data for ex-smokers taken from: Hrubek Z, McLaughlin J. Former cigarette smoking and mortality among U.S. veterans: A 26-year followup, 1954−1980. In: Burns D, Garfinkel L, Samet J, eds., Changes in cigarette-related disease risks and their implication for prevention and control. National Institutes of Health, 1997: p 509.

3. In the cases when the age categories in the railroad workers did not match the age groups from these two reports, the relative risks from the age categories that were most similar were used.

APPENDIX B: Confidence Intervals for Adjustment Methodology Proposed by Schlesselman and Axelson [Axelson, 1980; Schlesselman, 1978]

A generalized form of Axelson's formula appears below in equation (1). The terms in the numerator and the denominator of equation (1) consist of multiplying the proportions of individuals in each category of a confounder by the relative risk of disease associated with the confounding category. The numerator is the exposed group,

while the denominator is the unexposed group. Thus, the ratio of the numerator to the denominator defines S, which Schlesselman calls, ''the 'spurious' effect of C [a confounder] on the apparent relative risk under the assumption of no interaction'' (Schlesselman, p. 3). Other terms include k, which is the number of levels of the confounding variable. The first level of the confounder is the reference group with an implicit relative risk of disease equal to one. RR_i is the relative risk of disease due to the ith level of the confounder. P_{ci} is the proportion of individuals in the ith level of the confounder, noting that the numerator consists of only exposed individuals and the denominator consists of unexposed individuals. Therefore, x_1 can be called the weighted relative risk of disease due to the confounder for an exposed group and x_2 can be called the weighted relative risk of disease due to the confounder for an unexposed group.

$$S = \frac{\sum_{i=2}^{k} (RR_i) \hat{P}_{ci}, \text{exposed} + \left[1 - \sum_{i=2}^{k} \hat{P}_{ci}, \text{exposed} \right]}{\sum_{i=2}^{k} (RR_i) \hat{P}_{ci}, \text{unexposed} + \left[1 - \sum_{i=2}^{k} \hat{P}_{ci}, \text{unexposed} \right]}$$
$$= \frac{x_{\text{exposed}}}{x_{\text{unexposed}}} = \frac{x_1}{x_2} \quad (1)$$

Any deviance of S from one represents the positive or negative effects of the confounder. More specifically when S is greater than one, the confounding effect is positive. When S is less than one, the confounding effect is negative.

Adjusting an estimate of relative risk for confounding effects calls for equation (2), in which RR_{obs}

$$RR_{true} = \frac{RR_{obs}}{S} \quad (2)$$

represents the relative risk of disease due to an exposure that has not considered the confounder and RR_{true} represents the relative risk of disease due to an exposure that has adjusted for the effects of the confounder. Dividing the observed relative risk by the adjustment factor yields an estimate of the relative risk, that accounts for the confounder.

Because S is subject to sampling error, the following transformations are necessary to calculate confidence limits for RR_{true}:

$$\ln RR_{true} = \ln RR_{obs} - \ln S \quad (3)$$
$$\text{var}(\ln RR_{true}) = \text{var}(\ln RR_{obs}) + \text{var}(\ln S) \quad (4)$$

The general formula for 95% confidence limits is the following:

$$95\% \text{ CI for } \ln(RR_{true}) = \ln(\hat{RR}_{true}) \pm 1.96\sqrt{\text{var}(\ln \hat{RR}_{true})}$$
$$(5)$$

and requires variance estimates for S and the variance for the observed relative risk. The estimate for the observed relative risk can be taken from various statistical packages that provide multivariate output. The S component is the ratio of the weighted relative risks in the exposed and unexposed groups. Its variance can be determined by the following equation:

$$\text{var}\,(\ln S) = \text{var}\,(\ln x_1) + \text{var}\,(\ln x_2) \qquad (6)$$

The following approximation can be made by the delta method:

$$\text{for } x_1: \text{var}\,(\ln x_1) \cong \frac{1}{x_1^2}\,\text{var}\,(x_1) \qquad (7)$$

The variance for x_1 can be calculated by the following:

$$\text{var}(x_1) = \text{var}\left\{ \sum_{i=2}^{k}(\text{RR}_i)P_{ci} + \left[1 - \sum_{i=2}^{k}P_{ci} \right] \right\} \qquad (8)$$

$$= \text{var}\left\{ 1 - \sum_{i=2}^{k}(\text{RR}_i - 1)P_{ci} \right\} \qquad (8a)$$

$$= \sum_{i=2}^{k}(\text{RR}_i - 1)^2\,\text{var}\,P_{ci}$$

$$+ 2\sum_{i=2}^{k}\sum_{\substack{j=2 \\ i<j}}^{k}(\text{RR}_i - 1)(\text{RR}_j - 1)\,\text{cov}\left[P_{ci}, P_{cj} \right] \qquad (8b)$$

A similar formula holds for x_2.

Calculating the variance of x_1 and the variance of x_2 for equation (8) requires the following variance and covariance calculations, noting that each is calculated separately for the exposed and unexposed groups.

$$\hat{P}_{ci} = \frac{C_i}{N_{\text{total}}} \qquad (9)$$

$$\text{var}\hat{P}_{ci} = \frac{\hat{P}_{ci}\left[1 - \hat{P}_{ci} \right]}{N_{\text{total}}} \qquad (10)$$

$$\text{cov}\left[\hat{P}_{ci}, \hat{P}_{cj} \right] = -\frac{(\hat{P}_{ci}\hat{P}_{ci})}{N_{\text{total}}} \qquad (11)$$

Calculating the proportion of individuals in each level of the confounder requires dividing the number of people in each category (C_i) by the total number of individuals (N_{total}).

This methodology assumes that the point estimates of relative risks of disease associated with the confounder have no variability. In actuality, such relative risks are subject to error; therefore modification is required for equation (8) to incorporate the variability of the estimates. Such modification, even if the data were available, is likely to be small and may be excessive to add to indirectly adjusted estimates.

REFERENCE

Albanes D, Heinonen OP, Taylor PR, Virtamo J, Edwards BK, Rautalahti M, Hartman AM, Palmgren J, Freedman LS, Haapakoski J, Barrett MJ, Pietinen P, Malila N, Tala E, Liippo K, Salomaa ER, Tangrea JA, Teppo L, Askin FB, Taskinen E, Erozan Y, Greenwald P, Huttunen JK. 1996. Alpha-Tocopherol and beta-carotene supplements and lung cancer incidence in the alpha-tocopherol, beta-carotene cancer prevention study: effects of base-line characteristics and study compliance [see comments]. J Natl Cancer Inst 88: 1560–1570.

Axelson O. 1980. Aspects of confounding and effect modification in the assessment of occupational cancer risk. J Toxicol Environ Health 6:1127–1131.

Bandera EV, Freudenheim JL, Graham S, Marshall JR, Haughey BP, Swanson M, Brasure J, Wilkinson G. 1992. Alcohol consumption and lung cancer in white males. Cancer Causes and Control 3:361–369.

Blair A, Steenland K, Shy C, O'Berg M, Halperin W, Thomas T. 1988. Control of smoking in occupational epidemiologic studies: methods and needs. Am J Ind Med 13:3–4.

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

Burns D, Shanks T, Choi W, Thun M, Heath C, Garfinkel L. 1997. The American Cancer Society Cancer Prevention Study I: 12-Year followup of 1 million men and women. In: Burns D, Garfinkel L, Samet J, editors. Changes in cigarette-related disease risks and their implication for prevention and control. National Institutes of Health, pp 113–304.

California Environmental Protection Agency. Part B: Health risk assessment for diesel exhaust, as approved by the scientific review panel. 4-2-1998. California Environmental Protection Agency.

Crump KS. 1999. Lung cancer mortality and diesel exhaust: reanalysis of a retrospective cohort study of U.S. railroad workers. Inhal Toxicol 11:1–17.

Crump, KS, Lambert, T, and Chen, C. Assessment of risk from exposure to diesel engine emissions. U.S. EPA Contract 68-02-4601, Work Assignment #182. 1991. Louisiana, Clement International Corporation/U.S. EPA.

Emmelin A, Nystrom L, Wall S. 1993. Diesel exhaust exposure and smoking: a case-referent study of lung cancer among Swedish dock workers. Epidemiology 4:237–244.

Ferris B. 1978. Epidemiology standardization project (American Thoracic Society). Am Rev Respir Dis 118:1–120.

Fontham ET. 1997. Diet and lung cancer [editorial; comment]. Cancer Causes Control 8:819–820.

Garshick E, Schenker MB, Munoz A, Segal M, Smith TJ, Woskie SR, Hammond SK, Speizer FE. 1987a. A case-control study of lung cancer and diesel exhaust exposure in railroad workers. Am Rev Respir Dis 135:1242–1248.

Garshick E, Schenker MB, Munoz A, Segal M, Smith TJ, Woskie SR, Hammond SK, Speizer FE. 1988. A retrospective cohort study of lung cancer and diesel exhaust exposure in railroad workers. Am Rev Respir Dis 137:820–825.

Garshick E, Schenker MB, Woskie SR, Speizer FE. 1987b. Past exposure to asbestos among active railroad workers. Am J Ind Med 12:399–406.

Health Effects Institute. 1995. Diesel exhaust: a critical analysis of emissions, exposure, and health effects. Health Effects Institute.

Hrubek Z, McLaughlin J. 1997. Former cigarette smoking and mortality among U.S. veterans: A 26-year followup, 1954–1980. In: Burns D, Garfinkel L, Samet J, editors. Changes in cigarette-related disease risks and their implication for prevention and control. National Institutes of Health, pp 501–527.

Hyland A, Cummings KM, Lynn WR, Corle D, Giffen CA. 1997. Effect of proxy-reported smoking status on population estimates of smoking prevalence. Am J Epidemiol 145:746–751.

Knekt P, Heliovaara M, Rissanen A, Aromaa A, Seppanen R, Teppo L, Pukkala E. 1991. Leanness and lung-cancer risk. Int J Cancer 49:208–213.

Kolonel LN, Hirohata T, Nomura AM. 1977. Adequacy of survey data collected from substitute respondents. Am J Epidemiol 106:476–484.

Lerchen ML, Samet JM. 1986. An assessment of the validity of questionnaire responses provided by a surviving spouse. Am J Epidemiol 123:481–489.

McLaughlin JK, Dietz MS, Mehl ES, Blot WJ. 1987. Reliability of surrogate information on cigarette smoking by type of informant. Am J Epidemiol 126:144–146.

Microsoft Excel 97. 1997. Microsoft Corporation.

Omenn GS, Goodman GE, Thornquist MD, Balmes J, Cullen MR, Glass A, Keogh JP, Meyskens FLJ, Valanis B, Williams JHJ, Barnhart S, Cherniack MG, Brodkin CA, Hammar S. 1996. Risk factors for lung cancer and for intervention effects in CARET, the Beta-Carotene and Retinol Efficacy Trial [see comments]. J Natl Cancer Inst 88:1550–1559.

Preston D, Pierce D. EPICURE. 1.8. 1993. Seattle, HiroSoft International Corporation.

Rogot E, Reid DD. 1975. The validity of data from next-of-kin in studies of mortality among migrants. Int J Epidemiol 4:51–54.

SAS. (6.12). 1996. SAS Institute Inc. Cary, NC.

Schlesselman JJ. 1978. Assessing effects of confounding variables. Am J Epidemiol 108:3–8.

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.

Stellman SD, Boffetta P, Garfinkel L. 1988. Smoking habits of 800,000 American men and women in relation to their occupations. Am J Ind Med 13:43–58.

Veierod MB, Laake P, Thelle DS. 1997. Dietary fat intake and risk of lung cancer: a prospective study of 51,452 Norwegian men and women. Eur J Cancer Prev 6:540–549.

Woskie SR, Smith TJ, Hammond SK, Schenker MB, Garshick E, Speizer FE. 1988a. Estimation of the diesel exhaust exposures of railroad workers: I. Current exposures. Am J Ind Med 13:381–394.

Woskie SR, Smith TJ, Hammond SK, Schenker MB, Garshick E, Speizer FE. 1988b. Estimation of the diesel exhaust exposures of railroad workers: II. National and historical exposures. Am J Ind Med 13:395–404.