

Smoking Imputation and Lung Cancer in Railroad Workers Exposed to Diesel Exhaust

Eric Garshick, MD, MOH,^{1,2*} Francine Laden, ScD,^{2,3}
Jaime E. Hart, AB,^{2,3} Thomas J. Smith, PhD,³ and Bernard Rosner, PhD²

Background An association between diesel exhaust exposure and lung cancer mortality in a large retrospective cohort study of US railroad workers has previously been reported. However, specific information regarding cigarette smoking was unavailable.

Methods Birth cohort, age, job, and cause of death specific smoking histories from a companion case-control study were used to impute smoking behavior for 39,388 railroad workers who died 1959–1996. Mortality analyses incorporated the effect of smoking on lung cancer risk.

Results The smoking adjusted relative risk of lung cancer in railroad workers exposed to diesel exhaust compared to unexposed workers was 1.22 (95% CI = 1.12–1.32), and unadjusted for smoking the relative risk was 1.35 (95% CI = 1.24–1.46).

Conclusions These analyses illustrate the use of imputation in record-based occupational health studies to assess potential confounding due to smoking. In this cohort, small differences in smoking behavior between diesel exposed and unexposed workers did not explain the elevated lung cancer risk. *Am. J. Ind. Med.* 49:709–718, 2006.

© 2006 Wiley-Liss, Inc.

KEY WORDS: diesel exhaust; lung cancer; multiple imputation; smoking

INTRODUCTION

There is concern that diesel exhaust is a lung carcinogen. Lung cancer mortality in a large cohort of US railroad workers with long-term exposure was recently assessed. US Railroad Retirement Board (RRB) work history records were used to conduct a retrospective assessment of lung cancer

mortality between 1959 and 1996 in 54,973 workers, and a 40% (95% confidence interval (CI) = 30–51%) elevated lung cancer risk among those working in diesel exhaust exposed jobs, compared to those in unexposed jobs was observed [Garshick et al., 2004]. Using these historical work records allowed the efficient assessment of lung cancer risk due to long-term occupational exposure to diesel exhaust. However, as is common in retrospective occupational studies, individual level information on smoking, a potential confounder, was not available.

Although cigarette smoking causes lung cancer, the degree of confounding depends on the extent that smoking behavior differs between workers with and without diesel exhaust exposure. There are several strategies for minimizing and assessing the degree of potential confounding attributable to smoking in retrospective studies. One method is to exclusively study workers within a single industry and socioeconomic class. Since smoking behavior is a correlate of socioeconomic status and occupational category, the degree that smoking habits will vary among exposure

¹Pulmonary and Critical Care Medicine Section, Medical Service, VA Boston Healthcare System, MA

²Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

³Exposure, Epidemiology and Risk Program, Department of Environmental Health, Harvard School of Public Health, Boston, MA

Grant sponsor: NIOSH; Grant number: CCR115818; Grant sponsor: NCI; Grant number: CA79725.

*Correspondence to: Eric Garshick, Pulmonary and Critical Care Medicine Section, 1400 VFW Parkway, West Roxbury, MA 02132. E-mail: eric.garshick@med.va.gov

Accepted 4 May 2006
DOI 10.1002/ajim.20344. Published online in Wiley InterScience
(www.interscience.wiley.com)

categories is expected to be small [Lee et al., 2004]. To specifically assess the degree that smoking varies among exposure groups, a survey in a representative sample of workers can be conducted. The common method that specifically uses this survey information was first suggested by Schlesselman [1978] and Axelson [1980] to calculate smoking adjustment factors. In this method, the proportions of current and former smokers in each exposure category (diesel exposed/unexposed) are used to weight literature-based lung cancer risks due to smoking. Inherent in this method is the assumption that smoking intensity and duration varied similarly over time in each exposure category, and that differences in smoking behavior among job related exposure categories are similar in workers participating in the survey and workers in the retrospective study.

In this report the results of adapting an additional method, multiple imputation, are presented to assess the potential for confounding, and to incorporate the effect of smoking duration, intensity, and cessation into estimates of lung cancer risk. Although multiple imputation methods have been used to estimate the impact of various degrees of missing information, including smoking histories, in epidemiologic studies [Kmetz et al., 2002; Arnold and Kronmal, 2003; Mishra and Dobson, 2004], this methodology has not been widely applied. In particular, this method has not been used to simulate smoking behavior in retrospective occupational health studies. Smoking histories from an accompanying case-control study conducted in US railroad workers were used to provide age, birth cohort, job, and cause of death specific smoking information to impute smoking behavior in the retrospective railroad worker cohort and assess its impact on estimates lung cancer in diesel-exposed workers.

MATERIALS AND METHODS

Retrospective Cohort

The cohort has been described in detail previously [Garshick et al., 1988, 2004]. The US railroad industry changed from steam to diesel-powered locomotives starting primarily after World War II and through the late 1950s [US Department of Labor Bureau of Labor Statistics, 1972], and was 95% diesel by 1959. The US Railroad Retirement Board (RRB) has maintained computerized work records since 1959 for all railroad workers, including a yearly listing of job codes and months worked through retirement. Male workers in jobs with and without diesel exhaust exposure (see exposure assessment below), age 40–64 in 1959 with 10–20 years of prior railroad work were selected. Cause of death information from 1959 to 1996 was available from the National Death Index and from death certificates obtained from the RRB and state health departments. Since primary lung cancer (ICD9 162) is usually rapidly fatal following diagnosis with little recent improvement in survival, cases

were defined by the underlying cause of death or by lung cancer appearing elsewhere on the death certificate or NDI record. There were few non-white railroad workers included in the job categories that were selected and therefore analysis was limited to white males. There were 54,973 white male US railroad workers in the cohort, and through 1996 there were 43,593 deaths, including 4,351 lung cancer deaths.

Exposure Assessment

Between 1981 and 1983, an industrial hygiene survey was conducted to validate exposure assignments in the jobs selected for inclusion in the retrospective cohort [Woskie et al., 1988a,b]. The jobs included in the survey were two main occupational categories with diesel exhaust exposure as a result of work on operating trains, engineers (engineers and firemen), and conductors (conductors, brakemen, and hostlers), and an unexposed referent group (signal maintainers, and clerks, that included ticket agents, station agents, and other clerks). A shop group (shop supervisors, machinists, and electricians) was also included in the cohort. It was later determined that the shop job codes selected were not specific for locomotive shops which had been measured, but included other shops where there was no exposure to diesel exhaust, such as box car repair and dead repair and complete rebuilding of engines. As a result, workers with these job codes were considered as a separate group whose exposure was uncertain.

Concentrations of respirable particles were measured over a work shift and were used to characterize exposure. Cigarette smoke contributed to the respirable particles collected and nicotine in each sample was used to adjust for and remove the contribution of cigarette particulate [Woskie et al., 1988a,b]. The amount of particulate in the total due to diesel exhaust varied depending on proximity to sources of diesel exhaust. Mean respirable PM adjusted for cigarette PM for workers on operating trains, engineer and the conductor groups, were 71 and 89 $\mu\text{g}/\text{m}^3$, respectively. Workers without exposure were workers with clerical jobs (33 $\mu\text{g}/\text{m}^3$) and signal maintainers (58 $\mu\text{g}/\text{m}^3$). Since diesel locomotives first introduced in the late 1940s and throughout the 1950s were said to be “smokier” than locomotives introduced later and there were no exposure measurements available there was uncertainty in estimating historical exposures [Woskie et al., 1988a,b]. Therefore, as in previous reports of this cohort, survival analyses were conducted by comparing lung cancer risk between exposed and unexposed workers rather than specifically incorporating the PM exposure estimates.

Case-Control Study

The original railroad worker case-control study was designed as a matched case-control study of lung cancer and

diesel exhaust exposure [Garshick et al., 1987]. Between March 1, 1981 and February 28, 1982, there were 15,059 deaths among US railroad workers eligible for benefits. Death certificates were collected in 87% of the deaths. Lung cancer deaths were identified by death certificate in railroad workers born in 1900 or thereafter and matched on age and date of birth with up to two randomly selected control deaths who died within 30 days of the case, after excluding workers who died of an accidental cause or cancer. Two additional case series were identified that included other cancer deaths and deaths due to chronic respiratory diseases, for a total of 5,290 deaths. Efforts were made to obtain cigarette-smoking histories from next-of-kin of these deceased workers using mail questionnaires followed by a phone call. Questions about smoking included the age that the deceased first and last smoked cigarettes, and the average amount smoked daily. There were 4,119 persons (79%) with this information, and percentages were similar across the case and control series. Exposure to diesel exhaust was categorized using the exposure groups used in the retrospective cohort study. Workers in job codes not included in the retrospective cohort study were classified into exposure groups based on similarity in work locations and duties.

Smoking History Imputation

Smoking behavior in the US varies based on birth cohort and race [US Department of Health and Human Services, 1997]. Therefore, workers were identified in the case-control study who were in the same birth cohort, race, and occupational categories of workers in the retrospective cohort study. There were 2,470 white male workers in these categories with smoking history information available in the case-control dataset that included workers age 40–59 in 1959 (i.e., born between 1900 and 1919; Table I). As smoking histories were only available on deceased workers, the imputation of smoking behavior is limited to 39,388 workers (76% of all workers in the cohort ages 40–59 in 1959) who died through the end of follow-up in the retrospective cohort. Smoking history (age started, age stopped, and average number of cigarettes smoked daily) was assigned to each worker in the cohort with random selection from men in the case-control data of the same (a) age and birth cohort in 5-year groups (i.e., ages 40–44, 45–49, 50–54, 55–59 at study entry in 1959), (b) job category (engineer, conductor, shop, clerk, or signal maintainer groups), and (c) whether the subject died of lung cancer or another cause. Smoking histories were available from 626 workers who died of lung cancer and 1,844 deaths from other causes (480 other cancer, 906 cardiovascular causes, 302 chronic respiratory disease, and 156 other causes) and five data sets with imputed smoking information were created. The Brigham and Women's Hospital and VA Boston Healthcare System Institutional Review Boards approved the protocol.

Statistical Analysis

Proportional hazard analyses were used to assess lung cancer mortality in each dataset. Person-time was calculated from January 1, 1959 to the earlier of date of death or December 31, 1996. As in previous analyses [Garshick et al., 2004], to account for a healthy worker survivor effect, an effect where both survival and duration of work increase as workers leave the workplace due to illness or death [Arrighi and Hertz-Picciotto, 1993, 1994, 1995], time-varying variables for total years worked and for years off work (usually time after retirement) were included in survival models. Age was controlled by stratification in 1-year categories. Effect modification by age in 1959 was assessed by creating interaction terms of 5-year age group (40–44, 45–49, 50–54, and 55–59 years of age) and job category in 1959. It is unusual for railroad workers to change job categories, and job category in 1959 is highly predictive (approximate 97% or greater) of future work in that category [Garshick et al., 1988]. The association of lung cancer mortality with cumulative years of exposure in 5-year duration categories was assessed as a time-varying covariate, starting in 1959 in the combined engineer and conductor groups. An indicator variable was included to account for any work in a shop job code. We also constructed models where the exposure was lagged by excluding exposure in the last 5, 10, and 15 years.

Each worker's smoking behavior during the analysis was imputed in a time-dependent manner between 1959 and 1981 and allowed to vary based on age of smoking initiation and smoking cessation to account for the effect of age-related changes in smoking behavior. Because the case-control study provided smoking history information in 1981–1982, and there was no specific smoking information available, after 1981 smoking behavior was not allowed to vary in the regression models. Two smoking-adjusted models were considered, one with pack-years and years quit smoking, and the other with years of smoking, average daily consumption, and years quit smoking.

A full discussion of multiple imputation is beyond the scope of this report. However, in comparison to bootstrap and other Monte Carlo simulation methods where many simulations are required, Rubin and others [Rubin and Schenker, 1991; Rosner, 2006] have demonstrated that there is little increase in precision by performing more than five imputations. The methodology provided by Rubin and others was used to combine results from the imputations and to assess the relative efficiency of using five imputations rather than a larger number [Rubin and Schenker, 1991]. The between and within imputation variance and total variance was calculated, and used to calculate large sample 95% CIs for the mean of each regression parameter estimate from the five datasets.

RESULTS

Description of Case-control and Cohort Data

Smoking information available for each birth cohort (ages 40–44, 45–49, 50–54, 55–59 in 1959),

job category (engineer, conductor, shop, clerk, or signal maintainer), and cause of death (lung cancer, not lung cancer) (i.e., 40 specific combinations) from the case-control study and used to impute smoking behavior are presented in Table I. The distribution of workers in the cohort based on these same groupings is also presented.

TABLE I. Distribution of the Case-control and Cohort Datasets Used in the Smoking Simulation by Exposure Categories and Age at Baseline (in 1959)

	Age in 1959				Total
	40–44	45–49	50–54	55–59	
Case-control smoking history data					
Number with smoking history available	532	640	642	656	2,470
1959 job groups					
Unexposed	201	286	332	309	1,128
Clerks	108	137	164	173	582
Signal maintainers	93	149	168	136	546
Engineers ^a	102	110	78	87	377
Conductors ^b	167	154	106	127	554
Shop ^c	62	90	126	133	411
Lung cancer deaths					
Unexposed	38	80	79	75	272
Clerks	21	36	38	41	136
Signal maintainers	17	44	41	34	136
Engineers ^a	39	25	27	20	111
Conductors ^b	45	45	33	29	152
Shop ^c	13	27	21	30	91
Retrospective cohort					
Number in each age category	12,424	10,991	8,967	7,006	39,388
1959 Job groups					
Unexposed	2,885	2,415	2,355	1,924	9,589
Clerks	2,133	1,692	1,743	1,486	7,054
Signal maintainers	752	723	612	448	2,535
Engineers ^a	2,595	2,266	1,721	1,355	7,937
Conductors ^b	4,704	4,115	2,673	1,862	13,354
Shop ^c	2,240	2,195	2,218	1,855	8,508
Lung cancer deaths					
Unexposed	292	241	199	138	870
Clerks	210	160	154	101	625
Signal maintainers	82	81	45	37	245
Engineers ^a	328	257	189	113	887
Conductors ^b	556	475	267	173	1,471
Shop ^c	241	220	182	184	827
Retirement year					
Median	1976	1974	1970	1966	
Inter-quartile range	1970–1978	1969–1975	1966–1972	1964–1968	
Years of service mean (SD)	29.8 (6.9)	28.1 (5.9)	25.5 (5.1)	22.8 (4.2)	

^aEngineers, firemen.

^bConductors, brakemen, hostlers.

^cShop workers.

Imputed Smoking History and Imputation Efficiency

Percent current, former, and never smokers, and among smokers, cigarettes per day, years of smoking, and pack years obtained in each of the five imputations, is averaged for each birth cohort in 1959 and each job group (Table II). The variation in results across imputed data sets reflects the statistical uncertainty attributable to variation in the random assignment based on job group, birth cohort, and cause of death specific smoking behavior. Within each job category and age group, there was little variation in smoking behavior among simulations as demonstrated by the small standard

deviation (approximately 1% or less) in assigned smoking history categories. Depending on the specific regression term and smoking model, the relative efficiency in using five imputations to estimate the effect of smoking ranged from 86% to 94%, where relative efficiency is defined as the variance of the estimate of the regression parameter with an infinite number of imputations divided by the comparable variance with a finite (five) number of imputations [Rubin and Schenker, 1991; Yuan, 2000]. This suggests that additional efforts to impute smoking behavior using these data would not meaningfully influence the results.

The engineer and conductors groups had greater proportions of current smokers than clerks and signal

TABLE II. Baseline Mean Percent (Standard Deviation) Current, Former, and Never Smokers, and Among Smokers, Cigarettes Smoked Per Day, Years of Smoking and Pack Years in the Five Simulated Datasets by Age and Job Group at Entry

Age in 1959	Job group in 1959					
	Clerks	Signalman	Engineers	Conductors	Shop	All Jobs
40–44						
n	2,133	752	2,595	4,704	2,240	12,424
Current	78.3 (1.0)	77.6 (0.8)	85.8 (1.0)	83.3 (1.0)	79.9 (0.3)	82.0 (0.4)
Former	10.4 (0.5)	4.2 (0.6)	6.7 (0.5)	5.8 (0.3)	8.1 (0.2)	7.1 (0.1)
Never	11.3 (0.6)	18.2 (0.8)	7.6 (0.6)	10.9 (0.8)	12.1 (0.3)	10.9 (0.3)
Cigarettes/day	26.6 (0.2)	29.1 (0.1)	28.7 (0.3)	28.1 (0.2)	24.9 (0.1)	28.1 (0.2)
Years of smoking	24.2 (0.1)	25.2 (0.2)	25.4 (0.2)	25.7 (0.1)	25.0 (0.1)	25.2 (0.1)
Pack years	33.3 (0.3)	37.9 (0.4)	37.8 (0.6)	37.4 (0.3)	35.5 (0.7)	36.5 (0.3)
45–49						
n	1,692	723	2,266	4,115	2,195	10,991
Current	72.4 (1.5)	73.3 (0.8)	75.4 (0.4)	81.1 (0.7)	67.3 (2.0)	75.3 (0.7)
Former	13.1 (0.7)	8.4 (0.8)	9.6 (0.7)	3.9 (0.3)	11.6 (0.9)	8.3 (0.1)
Never	14.5 (1.3)	18.3 (1.2)	15.0 (0.7)	15.0 (0.8)	21.1 (1.1)	16.4 (0.6)
Cigarettes/day	26.0 (0.3)	26.0 (0.5)	26.4 (0.2)	27.5 (0.1)	22.9 (0.4)	26.3 (0.1)
Years of smoking	29.1 (0.2)	29.9 (0.2)	28.9 (0.1)	30.1 (0.1)	29.5 (0.2)	29.5 (0.1)
Pack years	38.7 (0.5)	40.1 (0.9)	38.7 (0.5)	43.2 (0.4)	35.2 (0.7)	39.8 (0.1)
50–54						
n	1,743	612	1,721	2,673	2,218	8,967
Current	65.0 (0.5)	65.5 (1.1)	77.6 (0.7)	70.1 (0.3)	57.6 (1.3)	67.1 (0.3)
Former	11.3 (0.7)	13.7 (0.8)	9.8 (0.5)	12.2 (0.2)	12.8 (0.4)	11.8 (0.3)
Never	23.6 (0.6)	20.8 (0.7)	12.7 (0.5)	17.7 (0.4)	29.6 (1.2)	21.0 (0.2)
Cigarettes/day	25.4 (0.3)	24.1 (0.4)	25.5 (0.1)	27.6 (0.4)	22.4 (0.3)	25.3 (0.1)
Years of smoking	33.3 (0.3)	33.1 (0.3)	34.7 (0.1)	33.1 (0.1)	32.6 (0.3)	33.5 (0.1)
Pack years	43.7 (0.8)	40.3 (0.7)	44.9 (0.2)	46.3 (0.2)	37.8 (0.7)	43.2 (0.2)
55–59						
n	1,486	448	1,355	1,862	1,855	7,006
Current	51.5 (0.9)	52.4 (0.5)	59.7 (1.2)	67.6 (0.4)	59.1 (1.4)	59.5 (0.5)
Former	23.8 (0.7)	20.4 (2.7)	24.8 (1.3)	12.3 (0.7)	11.4 (0.9)	17.4 (0.6)
Never	24.7 (1.4)	27.3 (2.2)	15.5 (0.7)	20.1 (0.8)	29.4 (1.5)	23.1 (0.6)
Cigarettes/day	22.4 (0.4)	24.5 (0.5)	21.8 (0.1)	27.4 (0.4)	25.9 (0.3)	24.6 (0.1)
Years of smoking	34.6 (0.1)	37.4 (0.5)	36.9 (0.1)	37.8 (0.2)	37.7 (0.2)	36.9 (0.1)
Pack years	40.8 (0.7)	47.4 (1.1)	41.8 (0.3)	51.0 (0.7)	49.7 (0.7)	47.0 (0.3)

maintainers and fewer never smokers for all age groups. There were small differences in average daily cigarette consumption, smoking duration, and pack years across job groups. In general, engineers and conductors had slightly more pack years of smoking than other workers. Although a greater proportion of younger workers at study entry in 1959 smoked, differences in smoking behavior between diesel exposed and unexposed job categories were less compared to older workers. For example, the proportion of current smokers in workers ages 40–44 at study entry varied from 77% to 78% in the clerks and signal maintainers to 85% in engineers and 83% in conductors. Among older workers age 55–59 at entry, the proportion of current smokers in the clerks and signal maintainers was approximately 50%, but was 59% among the engineers and 68% in the conductors.

Pack years were considered in four categories. Based on five imputations, the relative risk of lung cancer increased with the number of pack years (>0 to <25, RR = 3.61; 95% CI = 2.37–5.51; 25 to <50, RR = 6.44; 95% CI = 4.71–8.81; 50 to <75, RR = 8.62; 95% CI = 5.82–12.8; ≥75, RR = 10.1; 95% CI = 7.18–14.1, respectively) for persons

smoking within a year of death. In the same model, the reduction in risk associated with quitting smoking within 2–5 years of death was not statistically significant (RR = 0.94; 95% CI = 0.81–1.08), but for quitting smoking 6 or more years before death the RR was 0.70 (95% CI = 0.63–0.77). In additional models that included years of smoking, average daily consumption, and years quit smoking, lung cancer risk increased with smoking duration and average amount smoked, and decreased with smoking cessation (details not shown). When terms for diesel exhaust exposure were included in the models (as described below), the effect of cigarette smoking was similar to the unadjusted models.

Lung Cancer Mortality and Work in Diesel Exposed Jobs

As in previous analyses, workers in the engineer and conductor groups based on job in 1959 had an increased risk of lung cancer mortality, controlling for attained age, total years worked, and time since last worked (Table III). After adjustment for cigarette smoking, the risks among these

TABLE III. Relative Risk of Lung Cancer Mortality 1959–1996 by 5-Year Age Group and Job Title at Study Entry in 1959*

	Age group in 1959			
	40–44	45–49	50–54	55–59
Unexposed (Reference)				
Cases	292	241	199	138
Person years	71,714	56,703	49,135	34,902
RR	1.0	1.0	1.0	1.0
Engineer				
Cases	328	257	189	113
Person years	62,663	51,236	35,098	23,779
Smoking unadjusted RR ^a 95% CI	1.40 (1.19–1.64)	1.32 (1.11–1.58)	1.45 (1.19–1.77)	1.27 (0.99–1.63)
Smoking adjusted RR ^b 95% CI	1.27 (1.08–1.50)	1.26 (1.05–1.51)	1.22 (1.00–1.50)	1.09 (0.85–1.40)
Conductor				
Cases	556	475	267	173
Person years	114,897	93,642	54,577	32,369
Smoking unadjusted RR ^a 95% CI	1.25 (1.08–1.44)	1.29 (1.10–1.50)	1.29 (1.07–1.55)	1.39 (1.11–1.74)
Smoking adjusted RR ^b 95% CI	1.17 (1.01–1.36)	1.17 (1.00–1.37)	1.16 (0.97–1.41)	1.17 (0.92–1.49)
Shop worker				
Cases	241	220	182	184
Person years	55,093	51,004	46,763	34,546
Smoking unadjusted RR ^a 95% CI	1.07 (0.90–1.27)	1.02 (0.85–1.23)	0.94 (0.77–1.15)	1.32 (1.06–1.65)
Smoking adjusted RR ^b 95% CI	1.05 (0.88–1.25)	1.16 (0.96–1.41)	1.02 (0.82–1.26)	1.23 (0.97–1.55)
Engineer and conductor groups combined				
Cases	884	732	456	286
Person years	177,561	144,878	89,675	56,148
Smoking unadjusted RR ^a 95% CI	1.30 (1.14–1.49)	1.30 (1.12–1.50)	1.35 (1.14–1.59)	1.34 (1.09–1.64)
Smoking adjusted RR ^b 95% CI	1.21 (1.05–1.38)	1.20 (1.03–1.39)	1.19 (1.00–1.41)	1.14 (0.92–1.41)

*Interaction terms for individual job titles and age groups included in one model. A separate model with engineers and conductors combined is also presented.

^aAdjusted for attained age, years of employment and time off work as time dependent covariates.

^bSmoking adjusted models also include pack years and years quit smoking as time dependent covariates.

groups decreased but overall remained elevated. Similar results were obtained regardless of the specific smoking-related variables used to adjust for smoking (results for models that included years of smoking, average daily consumption, and years quit smoking not shown). After adjustment for smoking, there was more evidence of confounding by smoking among older workers ages 55–59 at study entry who were in the engineer group and conductor group (Table III) and for engineers age 50–54 than for younger workers. Among shop workers, the risks were not significantly elevated with the exception of workers aged 55–59 at study entry, and no consistently elevated risk was observed among shop workers after smoking adjustment.

The relationship of cumulative years of work in jobs with diesel exposure (engineer or conductor groups combined) and lung cancer risk was assessed in models without an exposure lag, and excluding exposure in the year of death and the preceding 4, 9, or 14 years (referred to as exposure lags of 5, 10, and 15 years). Lung cancer mortality risk was elevated in all exposure categories, but did not consistently increase with years of exposure after 1959. Results were similar regardless of the exposure lag model and are presented in Table IV for no lag and a 5-year lag. Adjustment for smoking attenuated the relative risks but did not change the pattern with increasing years of exposure. The smoking unadjusted relative risk (Table IV) for any diesel exposure (using a 5-year lag,) was 1.35 (95% CI = 1.24–1.46). The RR was attenuated to 1.22 (95% CI 1.12–1.32), after either adjusting for pack years and years quit smoking or including smoking duration, average daily consumption, and years quit smoking. In previous analyses [Garshick et al., 2004], exposure in the 5 years before death did not significantly contribute to mortality. Lung cancer mortality was also inversely related to total years worked, was greatest in the first years after leaving

work, and there was no significant effect modification based on diesel exposure on years off work (data not shown).

DISCUSSION

A retrospective assessment of lung cancer mortality over 38 years of follow-up was conducted in 39,388 deceased railroad workers aged 40–59 at entry (1959), using job, age, and birth cohort specific smoking histories imputed from a companion case-control study and allowed to vary in a time dependent manner. Disregarding exposure in the 5 years before death, the unadjusted relative risk for workers in jobs with any diesel exhaust exposure compared with workers without regular work in an exposed job was 1.35 (95% CI = 1.24–1.46). After smoking adjustment the excess risk was attenuated but remained significantly elevated (RR = 1.22; 95% CI = 1.12–1.32). There was no increase in risk with increasing years of exposure, a finding also noted in previous analyses of the entire cohort and without imputed smoking histories [Garshick et al., 2004]. Among older workers, adjustment for differences in smoking behavior resulted in a slightly greater reduction in risk than it did in younger workers. For example, based on results presented in Table III, the smoking unadjusted relative risk in the workers 55–59 at study entry in the engineer and conductor group combined was attenuated by a factor of 1.18 (ratio of smoking unadjusted/smoking adjusted relative risk). In contrast, in the combined engineer and conductor groups age 40–44 at study entry, the smoking unadjusted relative risk was attenuated by a smaller factor of 1.07. Whereas these differences based on birth cohort may be interpreted as small, they are also consistent with the greater differences in smoking behavior among job groups in older workers as demonstrated in Table II. These findings are consistent with

TABLE IV. Smoking Adjusted and Unadjusted Relative Risks of Lung Cancer Mortality Based on Either Any Exposure or Cumulative Years of Work in an Engineer or Conductor Job Group

		Any work		Years of work as engineer or conductor				
		Unexposed	Exposed	0 to <5 years	5 to <10 years	10 to <15 years	15 to <20 years	≥20 years
No lag	Cases	832	2,368	261	423	661	782	241
	Person years	205,938	469,755	120,110	125,465	115,834	89,637	18,709
Smoking unadjusted	RR ^a 95% CI	1.0	1.33 (1.23–1.44)	1.35 (1.15–1.57)	1.40 (1.23–1.59)	1.42 (1.27–1.58)	1.24 (1.11–1.38)	1.24 (1.05–1.46)
Smoking adjusted	RR ^b 95% CI	1.0	1.20 (1.11–1.30)	1.21 (1.04–1.42)	1.26 (1.14–1.42)	1.27 (1.14–1.42)	1.12 (1.01–1.25)	1.14 (0.97–1.36)
5-year lag ^c	Cases	895	2,305	330	449	615	707	204
	Person years	310,226	365,468	104,849	103,117	86,883	59,806	10,814
Smoking unadjusted	RR ^a 95% CI	1.0	1.35 (1.24–1.46)	1.44 (1.25–1.67)	1.36 (1.20–1.55)	1.36 (1.22–1.52)	1.28 (1.14–1.43)	1.32 (1.11–1.58)
Smoking adjusted	RR ^b 95% CI	1.0	1.22 (1.12–1.32)	1.31 (1.12–1.51)	1.23 (1.08–1.39)	1.23 (1.10–1.38)	1.16 (1.03–1.30)	1.22 (1.02–1.47)

^aAdjusted for age, years of employment, time-off work, and any work in a shop category as time-dependent covariates.

^bAdditionally adjusted for pack years and years quit smoking.

^cWork in an engineer or conductor job group in the year of death and 4 years before is not included as exposure.

the main results of the case-control study where younger workers who would have been age 42 or less in 1959 had an elevated risk of lung cancer that was similar with or without smoking adjustment [Garshick et al., 1987]. Overall, these results indicate that the observed elevated risk of lung cancer mortality in the diesel-exposed compared to unexposed workers cannot be completely explained by differences in smoking behavior.

In contrast to others conducting sensitivity analysis using externally obtained smoking information [Steenland and Greenland, 2004], an advantage of using data from the case-control study is that job and disease-specific smoking data are available. However, there are several potential limitations regarding the smoking history information used in the imputation. The smoking history information was not obtained directly from the worker, but was obtained from surrogate responders. However, as described previously, surrogates are able to accurately report smoking status and smoking duration. Although surrogates tend to over estimate rather than under report amount smoked [Rogot and Reid, 1975; Kolonel et al., 1977; Lerchen and Samet, 1986; McLaughlin et al., 1987; Hyland et al., 1997], it is unlikely that misclassification by a surrogate is likely to differ based on diesel exhaust exposure category.

An additional limitation is that since lung cancer cases and non-cases died within a 1-year period, they may not be representative of the smoking experience nor accurately reflect cause specific mortality of the entire cohort. The assignment of smoking histories based on a future cause of death (lung cancer) might also be questioned, but is justified since persons with lung cancer typically smoke more over a lifetime than persons without lung cancer. It was also not possible to condition the assignment of smoking histories on other specific causes of death since there were insufficient numbers when divided by birth cohort and job category. It is also possible that the smoking behavior of workers who died in 1981–1982 might not reflect the smoking histories of workers who died in the earlier years and later years of the cohort. However, the case-control study smoking history data was from workers who died at the approximate midpoint of the retrospective cohort study and who therefore are likely to have representative smoking histories.

Efforts were made to use smoking data that were representative of workers in the retrospective cohort study by selecting workers from the case-control database who were in the same birth cohort and age and year specific smoking behavior were calculated whenever possible during the imputation. In comparison to the imputed birth cohort specific smoking rates presented in Table II, the rates reported among US white males in 1959 available from National Health Interview Surveys (NHIS) [US Department of Health and Human Services, 1997] are slightly lower. Forever smokers ages 40–44, 45–49, 50–54, and 55–59, the NHIS rates are 82.1, 82.8, 80.6, and 77.9%, respectively, and

for current smokers are 70.2, 68.2, 63.0, and 57.3%, respectively. It is likely that the NHIS-based US rates are lower since the imputed rates are based on smoking histories obtained from deceased workers whose causes of death included smoking related causes. Although other birth cohort specific smoking information is not available, among 2,571 male US railroad workers ages 40–59 in 1957–1959 and who were enrolled in a study to assess cardiovascular health, only 59% were current cigarette smokers [Menotti et al., 2004]. Using occupation-specific information from the NHIS in 1978–1980 the prevalence of ever and current smoking among currently employed railroad workers was 68.5% and 44.3%, respectively [Brackbill et al., 1988]. In 703 rail conductors included in the survey, 61.6% were ever smokers and 40.7% were current smokers. Based on data from the American Cancer Society Prevention Study II in 1982 [Stellman et al., 1988] in a sample of 1,166 railroad workers, 33.6% were current cigarette, pipe, or cigar smokers, and 47% were former smokers. Overall, these data suggest that railroad workers in our cohort have historical smoking rates similar to US rates. In addition, although only deceased workers were included in the analysis, the overall effect of diesel exhaust exposure on lung cancer unadjusted for cigarette smoking was similar to the analysis using the entire cohort [Garshick et al., 2004].

Despite its limitations, the information available from the railroad worker case-control study is the most comprehensive database available describing job-specific smoking behavior among railroad workers [Garshick et al., 1987; Larkin et al., 2000]. As part of the original study design, railroad workers selected for inclusion in the cohort were likely to have similar smoking behaviors. This assumption was previously tested by using Schlesselman and Axelson methods to assess the distribution of job and birth cohort specific smoking habits [Larkin et al., 2000]. These smoking rates were used to weight literature-based lung cancer rates (diesel exposed/unexposed) to calculate smoking adjustment factors that generally ranged from 1.1 to 1.2. Using these factors a smoking-adjusted risk lung cancer of diesel exhaust exposure ranging from 1.17 to 1.27 in the full cohort was estimated [Garshick et al., 2004]. These estimates are similar to the results obtained using multiple imputation methods (smoking unadjusted RR = 1.35; 95% CI = 1.24–1.46; smoking adjusted RR = 1.22; 95% CI = 1.12–1.32). Despite these efforts, small relative risks may be influenced by residual confounding. This appears unlikely since adjustment for smoking using different methods provided similar results.

As in previous analyses in this cohort, lung cancer risk based on years of work in a diesel exposed job after 1959 did not increase [Garshick et al., 2004]. This association may be explained by a healthy worker survivor effect despite adjustment for employment status [Arrighi and Hertz-Picciotto, 1993, 1994, 1995]. Exposure to locomotives

during the 1950s and early 1960s in comparison to exposure to locomotives during other later periods would result [Liukonen et al., 2002; Verma and Finkelstein, 2002; Verma et al., 2003] in a temporal decrease in exposure intensity that would contribute to the lack of an exposure-response relationship.

To conclude, an application is illustrated where smoking histories available from a smaller sample of workers are used to impute smoking histories in a larger cohort where this information is not available. Smoking behavior was imputed using birth cohort, age, job specific smoking, and cause of death (lung cancer or not) specific information. The results indicate that small differences in smoking behavior between diesel exposed and unexposed workers does not explain the elevated lung cancer risk in the retrospective cohort, and are consistent with previous findings that adjust for potential confounding by smoking using other methods. This analysis demonstrates that it is possible to both consider potential confounding by smoking and take advantage of historical work records to identify a health risk in a timely and cost-effective manner when prospective data are not available.

ACKNOWLEDGMENTS

The authors thank Hongshu Guan for programming assistance; Emma Larkin and Stacey Campbell for data management; and the Railroad Retirement Board, in particular, Eileen Binkus and Anne Alden.

REFERENCES

- Arnold AM, Kronmal RA. 2003. Multiple imputation of baseline data in the cardiovascular health study. *Am J Epidemiol* 157:74–84.
- Arrighi HM, Hertz-Picciotto I. 1993. Definitions, sources, magnitude, effect modifiers, and strategies of reduction of the healthy worker effect. *J Occup Med* 35:890–892.
- Arrighi HM, Hertz-Picciotto I. 1994. The evolving concept of the healthy worker survivor effect. *Epidemiology* 5:189–196.
- Arrighi HM, Hertz-Picciotto I. 1995. Controlling for time-since-hire in occupational studies using internal comparisons and cumulative exposure. *Epidemiology* 6:415–418.
- Axelsson O. 1980. Aspects of confounding and effect modification in the assessment of occupational cancer risk. *J Toxicol Environ Health* 6:1127–1131.
- Brackbill R, Frazier T, Shilling S. 1988. Smoking characteristics of US workers, 1978–1980. *Am J Ind Med* 13:5–41.
- Garshick E, Schenker MB, Munoz A, Segal M, Smith TJ, Woskie SR, Hammond SK, Speizer FE. 1987. 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, Laden F, Hart JE, Rosner B, Smith TJ, Dockery DW, Speizer FE. 2004. Lung cancer in railroad workers exposed to diesel exhaust. *Environ Health Perspect* 112:1539–1543.
- 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.
- Kmetz A, Joseph L, Berger C, Tenenhouse A. 2002. Multiple imputation to account for missing data in a survey: Estimating the prevalence of osteoporosis. *Epidemiology* 13:437–444.
- Kolonel LN, Hirohata T, Nomura AM. 1977. Adequacy of survey data collected from substitute respondents. *Am J Epidemiol* 106:476–484.
- Larkin EK, Smith TJ, Stayner L, Rosner B, Speizer FE. 2000. Diesel exhaust and lung cancer: Adjustment for the effects of smoking in a retrospective cohort study. *Am J Ind Med* 38:399–409.
- Lee DJ, LeBlanc W, Fleming LE, Gomez-Marin O, Pitman T. 2004. Trends in US smoking rates in occupational groups: The National Health Interview Survey 1987–1994. *J Occup Environ Med* 46:538–548.
- Leichen ML, Samet JM. 1986. An assessment of the validity of questionnaire responses provided by a surviving spouse. *Am J Epidemiol* 123:481–489.
- Liukonen LR, Grogan JL, Myers W. 2002. Diesel particulate matter exposure to railroad train crews. *AIHA J (Fairfax, Va)* 63:610–616.
- 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.
- Menotti A, Kromhout D, Blackburn H, Jacobs D, Lanti M. 2004. Forty-year mortality from cardiovascular diseases and all causes of death in the US Railroad cohort of the Seven Countries Study. *Eur J Epidemiol* 19:417–424.
- Mishra GD, Dobson AJ. 2004. Multiple imputation for body mass index: Lessons from the Australian Longitudinal Study on Women's Health. *Stat Med* 23:3077–3087.
- 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.
- Rosner B. 2006. *Fundamentals of Biostatistics*. 6th edition. Belmont, CA: Duxbury.
- Rubin DB, Schenker N. 1991. Multiple imputation in health-care databases: An overview and some applications. *Stat Med* 10:585–598.
- Schlesselman JJ. 1978. Assessing effects of confounding variables. *Am J Epidemiol* 108:3–8.
- Steenland K, Greenland S. 2004. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *Am J Epidemiol* 160:384–392.
- 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.
- US Department of Health and Human Services. 1997. Changes in cigarette related disease risk and their implication for prevention and control. *Smoking and Tobacco Control Monograph 8*. Rockville, Maryland: US Department of Health and Human Services, Public Health Service, National Cancer Institute.
- US Department of Labor Bureau of Labor Statistics. 1972. *Railroad technology and manpower in the 1970s*. Washington, DC: USGPO. p 71.
- Verma DK, Finkelstein JN. 2002. Cancer risk from diesel exhaust exposure in the Canadian railroad industry: A feasibility study. In: Diesel Epidemiology Working Group editor. *Research Directions for Improve Estimates of Human Exposure and Risk from Diesel Exhaust*. Boston, MA: Health Effects Institute.

Verma DK, Finkelstein MM, Kurtz L, Smolyne K, Eyre S. 2003. Diesel exhaust exposure in the Canadian railroad work environment. *Appl Occup Environ Hyg* 18:25–34.

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.

Yuan Y. 2000. Multiple imputation for missing data: Concepts and new developments. SAS Institute Inc. <http://www.sas.com/rnd/app/papers/multipleimputation.pdf>.