ORIGINAL ARTICLE

Reducing healthy worker survivor bias by restricting date of hire in a cohort study of Vermont granite workers

Katie M Applebaum, Elizabeth J Malloy, Ellen A Eisen

Occup Environ Med 2007;**64**:681–687. doi: 10.1136/oem.2006.031369

See end of article for authors' affiliations

Correspondence to:
Dr E A Eisen, Harvard
School of Public Health,
Department of
Environmental Health, 665
Huntington Ave, Building I-
Room 1401, Boston, MA
02115, USA; eeisen@hsph.
harvard.edu

Accepted 30 March 2007
Published Online First
16 May 2007

Objective: To explore the healthy worker survivor effect (HWSE) in a study of Vermont granite workers by distinguishing "prevalent" from "incident" hires based on date of hire before or after the start of follow-up.

Methods: Records of workers between 1950 and 1982 were obtained from a medical surveillance programme. Proportional hazards models were used to model the association between silica exposure and lung cancer mortality, with penalised splines used to smooth the exposure-response relationship. A sensitivity analysis compared results between the original cohort and subcohorts defined by restricting date of hire to include varying proportions of prevalent hires.

Results: Restricting to incident hires reduced the 213 cases by 74% and decreased the exposure range. The maximum mortality rate ratio (MRR) was close to twofold in all subcohorts. However, the exposure at which the maximum MRR was achieved decreased from 4.0 to 0.6 mg-year/m³ as the proportion of prevalent hires decreased from 50% in the original cohort to 0% in the subcohort of incident hires.

Conclusion: Despite loss in power and restricted exposure range, decreasing the relative proportion of prevalent to incident hires reduced HWSE bias, resulting in stronger evidence for a dose-response between silica exposure and lung cancer mortality.

The healthy worker survivor effect (HWSE) remains a potential source of bias even in occupational epidemiological studies with an internal comparison group. In this paper, we focus on how a study's cohort definition can influence the HWSE and impact the exposure-response. We will do this with an example based on the Vermont granite workers.

The Vermont granite workers study has an important place in the literature involving silica and cancer.^{1–4} Earlier analyses of the mortality experience of the Vermont granite industry relied on an external comparison group and job titles as surrogates for silica exposure.^{1–2} In 2004, Attfield and Costello incorporated quantitative exposure information into models for lung cancer mortality.⁴ A categorical analysis revealed an increasing trend in the dose-response, except for a decline in the highest exposure category (≥ 6.0 mg-year/m³). The authors noted that 83% of those in the highest category of cumulative exposure had worked for at least 20 years by the start of follow-up, and that these self-selected survivors were likely to cause a healthy worker effect.

The HWSE results from less healthy workers transferring to jobs with lower exposure or leaving the workforce altogether. The healthier workers continue to be exposed (at higher levels), thus attaining higher cumulative exposures. As a result of this self-selection, the risk of disease at higher exposures appears to plateau or may even decrease. Sources of the HWSE bias and analytical approaches to reduce it in occupational epidemiology have been discussed previously.^{5–7}

Study design can impact the healthy worker effect. For example, cross-sectional studies of active workers over-sample longer term workers.⁸ Koskela *et al*⁹ used the term "cross-sectional follow-up" to describe follow-up studies of workers identified at one point in time. They found that these workers were healthier and had lower mortality rates than subjects captured as they enter the workforce (that is, open or inception cohort). Often, as is the case for the Vermont granite study, the cohort is a hybrid of these two approaches and includes both subjects already at work at the start of follow-up and others

who entered the cohort when they were hired. We were interested in how the distinction between these workers in the study population exaggerated (or minimised) the HWSE in the Vermont granite industry study and impacted the association between silica exposure and lung cancer mortality.

We distinguish two types of subjects, referred to as prevalent hires and incident hires. Those employees already at work at the start of the start of follow-up (that is, part of a cross-sectional follow-up identified at one point in time) will be referred to as "prevalent hires". "Incident hires" will describe those who first started working after the start of follow-up (that is, identified through the ongoing enrolment period of the inception cohort). The importance of this distinction is that to have made it into the Vermont granite workers study, prevalent hires must have endured the occupational exposure for some time and remained at work. Thus, prevalent hires may not be representative of all workers hired before the start of the study because those who were more susceptible to adverse health effects (either exposure related or unrelated) are more likely to have stopped working in the industry.¹⁰

The Vermont granite workers study provides an opportunity to examine the hypothesis that prevalent hires in the study population are a source of selection bias (HWSE) because more than half of the cohort was hired before the start of follow-up.

METHODS

Details on study design and exposure assessment have been described previously and are briefly reviewed here.^{1–4}

Cohort identification and follow-up

The most complete record of workers employed in the Vermont granite industry between 1950 and 1982 came from a medical surveillance programme. Granite workers who were x rayed at least once were eligible to be a part of this cohort study. A total of 5414 workers, all white males, were identified.

Abbreviations: HWSE, healthy worker survivor effect; MRR, mortality rate ratio

Information on vital status for 1950–82 was obtained from pension funds, the Vermont Division of Industrial Hygiene obituary notices of workers, and state vital records. For deceased workers, death certificates were obtained from the Quebec, New Hampshire, and Vermont Health Departments to determine the cause of death. When follow-up was extended to 1994, social security files on CD-rom were used to manually search for workers by last name to identify more deaths. A qualified nosologist coded the cause of death according to the Eighth Revision of the International Classification of Diseases. Study protocol was approved by the human subjects review committee at the Harvard School of Public Health.

Exposure assessment

Through their participation in the medical surveillance programme, granite workers provided work history information. Exposure assessment has been described in Attfield and Costello.⁴ Briefly, each worker’s quantitative silica exposure was estimated based on environmental surveys conducted between 1924 and 1977. Measured respirable silica dust concentrations for each job provided the basis for the exposure estimates. Cumulative exposure to respirable silica dust (mg-year/m³) was calculated for each granite worker by matching his work history with the estimated exposure for each job and year and summing the exposures over time. To be consistent with previous studies of lung cancer mortality and occupational exposures, we lagged exposure by 15 years.^{4 11–16}

Classification of subjects by date of hire

The Vermont granite cohort is comprised of both incident hire and prevalent hire subjects. A schematic for potential subjects in the cohort is presented in figure 1. Incident hires—that is, those hired during the follow-up period (starting in 1950 and later)—were captured as they entered the workforce. Prevalent hires worked before the start of follow-up and continued working until the start of follow-up in 1950. Those who started and stopped working before 1950 will be missing. Thus, the probability of being missing is related to length of employment as well as to health status.

Data analysis

Cox proportional hazards regression was used to estimate the mortality rate ratio (MRR) and 95% confidence interval (CI) for the relationship between silica and lung cancer mortality (SAS version 9.1). Risk sets were generated using age as the timeline. We controlled for calendar year achieved risk age because lung cancer mortality was increasing over the time period of this

study and silica dust exposures were decreasing. Calendar year was included as an ordinal variable representing increasing decade at risk age. We also adjusted for age at hire with a continuous variable. Smoking information was not available.

Cumulative exposures typically have non-linear relationships with occupational disease—for example, a plateau or decrease in risk among those with the highest exposures. Such patterns are generally attributed to the healthy worker effect.^{6 7} To avoid a linearity assumption, we initially analysed cumulative silica exposure (15-year lag) as a categorical variable. Quartiles based on the exposure distribution of the cases in the original cohort were used to define the categories. To compare the silica dose-response by hire status, the same categories for silica were also analysed after stratifying by incident or prevalent hire status. We tested for statistical interaction by comparing the log likelihood for the model with silica exposure, hire status, and their cross products to the log likelihood from a model without the cross products. However, there were differences in the distribution of exposure for incident and prevalent hires. Therefore, we also determined categories separately by hire status and analysed these for an association with lung cancer mortality. In categorical analyses stratified by hire status, cut-points were determined by the distribution of the cases in each stratum. A trend test was conducted for categorical analyses by including an ordinal term representing the categories with increasing cumulative silica exposure. All statistical tests were two-sided.

To more fully characterise the dose-response relationship for lung cancer mortality and silica, semi-parametric models based on penalised splines were fit using the approach suggested by Thurston, *et al.*¹⁷ We used penalised splines to allow for non-linearity because they are insensitive to the location or number of knots and have been applied in Cox models of occupational cohorts.^{16–18} Spline models were generated using R 1.9.0. Currently, R is unable to generate a penalised spline for an interaction term in a Cox model. Therefore, penalised spline models were fit to prevalent and incident hires in a stratified analysis. Akaike’s Information Criterion (AIC) was used to guide the selection of the optimal number of degrees of freedom.¹⁹ A sensitivity analysis examined the impact of moving the date of

Figure 1 Schematic of Vermont granite workers cohort definition. The circles with lines extending to the right represent possible person-time experiences of Vermont granite workers, which ended either as a lung cancer death (x) or being censored (|). Study follow-up began in 1950 and ended in 1994. The original design captured subjects who had worked anytime in between, regardless of date of hire.

Table 1 Selected characteristics for the Vermont granite workers study, 1950–94: comparison of prevalent and incident hires

	Prevalent hires* (n = 2603)	Incident hires† (n = 2578)
	Mean‡ (SD)	Mean (SD)
Lung cancer deaths		
Age at death	67.2 (8.8)	61.7 (9.6)
Time-independent		
Year of hire	1932 (13)	1961 (7)
Age at hire	24.2 (7.7)	28.1 (9.4)
Time-varying§		
Duration of follow-up	28.3 (13.8)	31.5 (8.0)
Duration of work	30.9 (13.1)	10.5 (9.4)
Time since hire	46.0 (11.5)	31.5 (8.0)
Calendar year at risk	1977 (10)	1984 (9)
Cumulative silica		
15-year lag, mg-year/m ³	3.7 (4.6)	0.3 (0.4)
Range	(0 to 48.2)	(0 to 2.0)

SD, standard deviation.
*Subjects who were already at work prior to the start of follow-up on 1 January 1950 and still at work on that date, includes 157 cases.
†Subjects who started working in the granite industry between 1 January 1950 (the start of follow-up) and 31 December 1982; includes 56 cases.
‡All mean differences between prevalent and incident hires were statistically significant (p<0.001).
§Time-varying variables evaluated for each subject at the end of follow-up.

hire relative to the start of follow-up. Exposure-response splines were generated for three cohort definitions: workers hired in 1940 or later (10 years before start of follow-up), hired in 1930 or later (20 years before start of follow-up), and workers hired in 1920 or later (30 years before start of follow-up).

RESULTS

A total of 5414 subjects were initially identified as having worked anytime between 1950 and 1982 and having at least one x ray. This surveillance programme is reported to have included 98% of the eligible workforce.²⁰ Thus, 98% of workers hired from 1950 and later (the incident hires) were captured. However, there is no complete record of all workers who started and stopped work before 1950; therefore, we cannot say what proportion of people who worked before 1950 was represented by the prevalent hires in this study.

Six subjects were excluded due to incomplete demographic data (for example, date of birth). There were 213 lung cancer deaths among the remaining 5408 Vermont granite workers in the original cohort. Upon determination of risk sets for cases, 227 subjects did not contribute to any risk set either because they were censored before the youngest case, or they were older than the oldest risk age at the start of the follow-up period. A total of 5181 subjects were included in the risk sets, and 52% of these were prevalent hires.

Table 1 compares descriptive characteristics for prevalent and incident hires, which together comprise the original data set. Overall, subjects were approximately evenly divided between prevalent hires (n = 2603) and incident hires (n = 2578) while 74% of the lung cancer deaths were prevalent hires (prevalent

Figure 2 Mortality rate ratio for cumulative silica exposure and lung cancer using the original cohort definition for the Vermont granite workers study. Mortality rate ratio estimated by a Cox proportional hazards model with penalised splines. Model used age as the timeline and controlled for calendar year and age at hire. The original cohort was comprised of subjects who worked in the Vermont granite industry anytime between the start of follow-up on 1 January 1950 and 31 December 1982, regardless of date of hire (for example, includes prevalent and incident hires).

Table 2 Association between lung cancer mortality and categorical cumulative silica in the Vermont granite workers study, 1950–94

Original cohort‡	Cumulative silica*	Cases n = 213	Mortality rate ratio† (95% CI)	p for trend
	0 to 0.4	53 (24.9%)	Referent	
	>0.4 to 1.3	53 (24.9%)	1.84 (1.22 to 2.77)	0.001
	>1.3 to 2.9	53 (24.9%)	2.54 (1.58 to 4.06)	
	>2.9 to 48.2	54 (25.4%)	2.35 (1.36 to 4.07)	
Stratified by hire status: cutpoints same as original cohort				
Prevalent hires§		n = 157		
		15 (9.6%)	Referent	
		37 (23.6%)	1.63 (0.89 to 2.98)	0.02
		51 (32.5%)	2.14 (1.18 to 3.90)	
Incident hires¶		n = 56		
		38 (67.9%)	Referent	
		16 (28.6%)	2.01 (1.05 to 3.83)	0.04
		2 (3.6%)	2.22 (0.49 to 10.02)	
		0	NA	
Stratified by hire status: cutpoints determined within hire status				
Prevalent hires§		n = 157		
	0 to 1.0	39 (24.8%)	Referent	
	>1.0 to 2.0	39 (24.8%)	1.29 (0.81 to 2.04)	0.15
	>2.0 to 3.6	39 (24.8%)	1.90 (1.18 to 3.08)	
Incident hires¶		n = 56		
	0 to 0.04	19 (33.9%)	Referent	
	>0.04 to 0.4	18 (32.1%)	1.39 (0.67 to 2.85)	0.02
	>0.4 to 2.0	19 (33.9%)	2.51 (1.14 to 5.55)	

*15-year lag, mg-year/m³.
†Models used age as the timeline and controlled for calendar year and age at hire.
‡Subjects who worked in granite industry anytime between 1 January 1950 (start of follow-up) and 31 December 1982 (includes incident and prevalent hires).
§Subjects who were already at work prior to the start of follow-up on 1 January 1950 and still at work on that date.
¶Subjects who started working in the granite industry between 1 January 1950 (the start of follow-up) and 31 December 1982.

Figure 3 Mortality rate ratio for cumulative silica exposure and lung cancer among only the prevalent hires in the Vermont granite workers study. Mortality rate ratio was estimated by a Cox proportional hazards model with penalised splines. Model used age as the timeline and controlled for calendar year and age at hire. Prevalent hires were already at work prior to the start of follow-up on 1 January 1950 and were still at work on that date.

hire cases = 157, incident hire cases = 56). On average, there was a 30-year difference in year of hire for prevalent and incident hires (1932 compared to 1961, respectively). Incident hires started working in the industry at a slightly older age than the prevalent hires and worked about 20 years less by the end of follow-up than prevalent hires. The duration of follow-up was not very different between the two groups, which shared the earliest possible date for the start of follow-up of 1 January 1950.

The most notable difference between prevalent and incident hires was the accumulation of exposure. Overall, prevalent hires had a higher mean cumulative silica exposure than the incident hires. This was due in part to their longer work tenure, but it was also influenced by the implementation of industry-wide control technologies in the 1940s, which reduced silica dust concentrations. Among incident hires, there was no difference in mean silica exposure between cases and non-cases (both 0.3 mg-year/m³). Among prevalent hires, however, the mean silica exposure was higher among non-cases (3.7 mg-year/m³) than cases (2.9 mg-year/m³).

The categorical models for cumulative exposure are presented in table 2. In the model for the original cohort, there was an elevated mortality rate with higher cumulative exposure, achieving more than a doubling in the mortality rate (MRR 2.54, 95% CI 1.58 to 4.06, *p* for trend = 0.001) before the dose-response began to decline. Stratifying by hire status, the mortality rate for prevalent hires increased to approximately a doubling relative to the lowest exposed category (MRR 2.14, 95% CI 1.18 to 3.90) and there was a statistically significant trend (*p* = 0.02). However, for the incident hires, 68% of cases fell within the first quartile of exposure of the original cohort, only two (4%) in the third quartile, and none in the highest. Although a test for interaction can be conducted (results were not significant), it was not informative of a difference in the dose-response because there was so little overlap in the exposure distributions of the prevalent and incident hires. When including prevalent and incident hires in the same categorical model, the incident hires therefore served as the

Figure 4 Mortality rate ratio for cumulative silica exposure and lung cancer among only the incident hires in the Vermont granite workers study. Mortality rate ratio was estimated by a Cox proportional hazards model with penalised splines. Model used age as the timeline and controlled for calendar year and age at hire. Incident hires started working in the Vermont granite industry between 1 January 1950 (the start of follow-up) and 31 December 1982.

referent group for the prevalent hires (71.7% of person-time in the referent group was from the incident hires).

Table 2 also presents stratified categorical analyses with the cutpoints for the quartiles determined separately for incident and prevalent hires. For prevalent hires, the mortality rate ratio increased to 1.90 (95% CI 1.18 to 3.08) and then declined in the highest exposure category. Because the incident hires had lower power due to the number of cases and a narrower exposure range, we analysed tertiles of exposure. For incident hires, we observed a dose response which reached a mortality rate ratio of 2.51 (95% CI 1.14 to 5.55) with a statistically significant trend (*p* = 0.02) over the low exposure range (0.4 to 2.0 mg-year/m³).

Penalised splines were fit and predicted MRR plotted against the range of exposure (figs 2 through 4). For the original cohort (fig 2 and table 3), the MRR increased linearly up to a

Table 3 Mortality rate ratios from Cox proportional hazards models with penalised spline models for silica and lung cancer in the Vermont granite study, 1950–94

Cohort definition	Maximum MRR* (95% CI)	Exposure† at maximum MRR	% Prevalent hires
Original cohort‡	2.50 (1.81 to 3.47)	4.0	50.2
Prevalent hires§	1.97 (1.52 to 2.56)	4.1	100.0
Incident hires¶	2.19** (1.27 to 3.76)	0.6	0.0

MRR, mortality rate ratio.
*Models used age as the timeline and controlled for calendar year and age at hire.
†Cumulative exposure with 15-year lag, in mg-year/m³.
‡Subjects who worked in granite industry anytime between 1 January 1950 (start of follow-up) and 31 December 1982 (includes prevalent hires and incident hires).
§Subjects who were already at work prior to the start of follow-up on 1 January 1950 and still at work on that date.
¶Subjects who started working in granite industry between 1 January 1950 (start of follow-up) and 31 December 1982.
**For incident hires, MRR of 14.89 (95% CI 1.63 to 135.76) was observed at 2.0 mg-year/m³.

Table 4 Mortality rate ratios from Cox proportional hazards models with penalised spline models for silica and lung cancer: comparison of cohort definitions for the Vermont granite study, 1950–94

Cohort definition	Maximum MRR* (95% CI)	Exposure† at maximum MRR	% Prevalent hires
Hired 1940 or later	2.60 (1.54 to 4.39)	2.25	27.2
Hired 1930 or later	2.45 (1.53 to 3.94)	3.51	36.1
Hired 1920 or later	2.29 (1.62 to 3.23)	4.55	45.7

*Models used age as the timeline and controlled for calendar year and age at hire.
†Cumulative exposure with 15-year lag, in mg-year/m³.

cumulative exposure of 4.0 mg-year/m³, where the MRR reached a maximum of 2.50 (95% CI 1.81 to 3.47) and then the MRR declined below the null. For prevalent hires (fig 3 and table 3), the shape of the dose-response curve was similar to the original cohort, although the maximum MRR was lower (1.97, 95% CI 1.52 to 2.56). Restricting to incident hires (fig 4 and table 3), the MRR increased up to 2.19 (95% CI 1.27 to 3.76) at a much lower cumulative exposure of 0.6 mg-year/m³. This was followed by a drop in the MRR, which increased dramatically, albeit with very wide confidence intervals, at the highest exposure of 2.0 mg-year/m³.

To further examine the impact of including prevalent hires, we conducted a sensitivity analysis. Figure 5 and table 4 present what occurred when an increasing proportion of prevalent hires was included in the analysis. Penalised splines containing the incident hires with prevalent hires hired within 10 years before the start of follow-up (hired 1940 or later), 20 years before the start of follow-up (hired 1930 or later), and hired as much as 30 years before the start of follow-up (hired 1920 or later). The steepest slope was among those hired 1940 and later. As the date of hire for cohort eligibility was moved further away from the start of follow-up (for example, earlier calendar date) and allowed for a higher proportion of prevalent hires, the exposure range increased and became more skewed (fig 5).

In table 4, as the percentage of prevalent hires decreased, the maximum MRR occurred at a lower cumulative exposure (from 4.55 mg-year/m³ for those hired 1920 or later (45.7% prevalent hires) to 2.25 mg-year/m³ among those hired 1940 or later (27.2% prevalent hires)). Furthermore, the maximum MRR was somewhat higher with fewer prevalent hires, although the lower limits of the confidence intervals were similar.

DISCUSSION

Prevalent hires were older at time of lung cancer death, younger at hire, and worked longer which allowed them to accumulate higher silica exposures than the incident hires. With penalised splines, we observed that the exposure-response increased and then declined in a non-linear fashion, whether incident hires and prevalent hires were combined or analysed separately. The main limitations to restricting the study to incident hires were the loss in power as the number of cases decreased by 74% and the dramatically reduced exposure range. A sensitivity analysis indicated that as the date of hire criterion moved closer to the start of follow-up, the exposure-response slope became steeper, allowing for a doubling in the mortality rate ratio to be reached at a lower cumulative exposure.

Two recent analyses of the Vermont granite industry cohort that included the extended follow-up period reached opposing conclusions about whether there was an association between lung cancer mortality and silica dust exposure.^{3 4 21} We demonstrated that HWSE bias may be exaggerated in the original cohort, which was used in both of the previous

Figure 5 Comparison of dose-response curves for silica exposure and lung cancer mortality from the Vermont granite industry defined by a different date of hire eligibility: 1940 or later, 1930 or later, and 1920 or later. Follow-up for all subjects began in 1950 regardless of date of hire. Mortality rate ratios estimated with Cox proportional hazards models using penalised splines. Models used age as the timeline and controlled for calendar year and age at hire.

analyses, with the presence of prevalent hires potentially weakening the association between silica and lung cancer. When we accounted for this selection bias, the lung cancer mortality rate increased with increasing cumulative silica up to 2.25 mg-year/m³, even when including subjects who started working up to 10 years before start of follow-up. Indeed, if we used the strictest cohort definition limited to incident hires only, all of whom started working after dust control measures had been implemented, there was evidence of an exposure-response relationship, exhibited with the steepest slope and statistically significant association between silica and lung cancer mortality when cumulative exposure reached 0.6 mg-year/m³.

These findings are consistent with other occupational studies of lung cancer where a subset of incident hires was examined.^{22 23} For example, in a study of a metal production industry in Colorado, the strongest standardised mortality ratio for lung cancer mortality was achieved at a lower cumulative cadmium exposure for incident hires than for the prevalent hires.²² In another study of Tacoma smelter workers, only incident hires exhibited a linear dose-response with increasing arsenic exposure, resulting in a stronger dose-response at lower levels of exposure.²³

Although HWSE will be minimised in a cohort composed of incident hires only, it is important to recognise that this bias will continue to be present because those who are less healthy will still be more likely to remove themselves from higher exposure earlier than those who are healthier. The use of lagging exposure has been suggested as a means of reducing HWSE.²⁴ In addition to the 15-year lag, we also examined models without a lag and with shorter lag periods (5 and 10 years), and the same pattern of stronger dose-response for incident hires persisted. Adjusting for time since hire or employment status has also been promoted as a means of controlling for this ongoing HWSE. However, there was no change in the estimates when time since hire or employment status was included in the model. Further, controlling for these

Main messages

- In workplace-based studies, *prevalent hires* are subjects who were already employed at the start of follow-up; *incident hires* are subjects who were hired after the start of follow-up.
- Including prevalent hires introduces bias in occupational cohort studies.
- Epidemiologists should minimise the proportion of prevalent hires included in occupational study populations.
- In studies that include prevalent hires, investigators should report the proportion and the distribution of hire dates so that potential bias can be assessed.

Policy implication

- Health standards derived from epidemiological evidence based on prevalent hires may not be adequately protective for workers.

time-varying factors, which in addition to their confounding role may also be intermediate variables, would induce bias.^{25–27} As a result, structural nested models have been proposed to address HWSE.^{28–29} A greater discussion of methods that address the ongoing HWSE in occupational cohorts is beyond the scope of this analysis.

An alternative explanation for the observed differences in the exposure-response relationships between prevalent and incident hires is measurement error. The cumulative exposure estimates were derived from six environmental surveys conducted between 1924 and 1977. It could be argued that exposure measurements improved with time, thereby reducing exposure misclassification and its associated bias towards the null. As a result, the stronger association with fewer prevalent hires could be an artifact of less measurement error. However, even if measurement error decreased with time, the bias induced by including prevalent hires would still be present. Furthermore, to the extent that measurement error decreased over time, focusing on incident hires will reduce error.

Incident hires were older, on average, than prevalent hires at the time of first employment. This difference was larger for cases than non-cases. Consider two subjects, both hired at age 25. The prevalent hire started work in 1940 and the incident hire started in 1960. The prevalent hire will be 20 years older than the incident hire in any given year of follow-up. If both subjects eventually die of lung cancer at age 65, only the prevalent hire will be identified as a case during follow-up (1950–94). As age is a risk factor for lung cancer, incident hires who start work older and die from lung cancer are more likely to be identified as cases than incident hires who start work younger.³⁰ To address potential confounding by age at hire, we adjusted for it in the models.

In this analysis, we could not control for smoking. The absence of smoking information is a common challenge for occupational studies based on historical data. It has been demonstrated that the magnitude of uncontrolled confounding due to smoking is modest in occupational studies of lung cancer.^{31–34} Most of this attention has focused on smoking as an unmeasured positive confounder, including discussion in the original analysis of this cohort.⁴ If, however, smoking cases are

more likely to leave work than non-smoking cases, smoking in fact could be operating as a negative confounder in this study.

The impact of healthy worker bias in studies of respiratory cancer has been questioned as symptoms do not typically appear until shortly before diagnosis.³⁵ Some have argued that the bias does exist and provided empirical evidence.^{6–24–36} When the exposure suspected of causing lung cancer is known to cause other adverse respiratory effects, which are, in turn, associated with lung cancer, then HWSE becomes even more plausible. In the case of silica, the association with lung cancer has been found to be stronger among silicotics, suggesting that silicosis may be on the causal pathway to lung cancer.³⁷ In addition, silica causes deficits in pulmonary function and x ray changes consistent with silicosis, which can lead to difficulty breathing, persistent cough, and bronchitis.³⁸ If these earlier health effects are related to silicosis, then they can bias associations with lung cancer because leaving work is related to both lower exposure (truncated exposures) and outcome.

We conclude that HWSE bias may be reduced in occupational cohort studies by restricting date of hire to be close to the start of follow-up. Further, we urge that authors of occupational epidemiological studies clearly state what the dates of hire and follow-up were for study subjects to ensure that this bias is limited in any analysis.

ACKNOWLEDGEMENTS

The authors express their appreciation to Michael Attfield for making the data available. This work was supported by grants T32 ES07155 and CA81345 from the National Institutes of Health.

Authors' affiliations

K M Applebaum, E A Eisen, Harvard School of Public Health, Department of Environmental Health, Boston, Massachusetts, USA

K M Applebaum, Harvard School of Public Health, Department of Genetics and Complex Diseases, Boston, Massachusetts, USA

E J Malloy, American University, Department of Mathematics and Statistics, Washington, DC, USA

Competing interests: None declared.

REFERENCES

- 1 **Davis LK**, Wegman DH, Monson RR, *et al.* Mortality experience of Vermont granite workers. *Am J Ind Med* 1983;**4**:705–23.
- 2 **Costello J**, Graham WG. Vermont granite workers' mortality study. *Am J Ind Med* 1988;**13**:483–97.
- 3 **Graham WG**, Costello J, Vacek PM. Vermont granite mortality study: an update with an emphasis on lung cancer. *J Occup Environ Med* 2004;**46**:459–66.
- 4 **Attfield MD**, Costello J. Quantitative exposure-response for silica dust and lung cancer in Vermont granite workers. *Am J Ind Med* 2004;**45**:129–38.
- 5 **Fox AJ**, Collier PF. Low mortality rates in industrial cohort studies due to selection for work and survival in the industry. *Br J Prev Soc Med* 1976;**30**:225–30.
- 6 **Arrighi HM**, Hertz-Picciotto I. Controlling the healthy worker survivor effect: an example of arsenic exposure and respiratory cancer. *Occup Environ Med* 1996;**53**:455–62.
- 7 **Stayner L**, Steenland K, Dosemeci M, *et al.* Attenuation of exposure-response curves in occupational cohort studies at high exposure levels. *Scand J Work Environ Health* 2003;**29**:317–24.
- 8 **Rothman KJ**, Greenland S. *Modern epidemiology*. Second edition. Philadelphia: Lippincott-Raven Publishers, 1998.
- 9 **Koskela RS**, Jarvinen E, Kolari PJ. Effect of cohort definition and follow-up length on occupational mortality rates. *Scand J Work Environ Health* 1984;**10**:311–16.
- 10 **Richardson D**, Wing S, Steenland K, *et al.* Time-related aspects of the healthy worker survivor effect. *Ann Epidemiol* 2004;**14**:633–9.
- 11 **Kauppinen T**, Heikkilä P, Partanen T, *et al.* Mortality and cancer incidence of workers in Finnish road paving companies. *Am J Ind Med* 2003;**43**:49–57.
- 12 **Boffetta P**, Burstyn I, Partanen T, *et al.* Cancer mortality among European asphalt workers: an international epidemiological study. II. Exposure to bitumen fume and other agents. *Am J Ind Med* 2003;**43**:28–39.
- 13 **Steenland K**, Mannetje A, Boffetta P, *et al.* Pooled exposure-response analyses and risk assessment for lung cancer in 10 cohorts of silica-exposed workers: an IARC multicentre study. *Cancer Causes Control* 2001;**12**:773–84.
- 14 **Steenland K**, Sanderson W. Lung cancer among industrial sand workers exposed to crystalline silica. *Am J Epidemiol* 2001;**153**:695–703.

- 15 **Checkoway H**, Heyer NJ, Seixas NS, *et al.* Dose-response associations of silica with nonmalignant respiratory disease and lung cancer mortality in the diatomaceous earth industry. *Am J Epidemiol* 1997;**145**:680–8.
- 16 **Eisen EA**, Agalliu I, Thurston SW, *et al.* Smoothing in occupational cohort studies: an illustration based on penalised splines. *Occup Environ Med* 2004;**61**:854–60.
- 17 **Thurston SW**, Eisen EA, Schwartz J. Smoothing in survival models: an application to workers exposed to metalworking fluids. *Epidemiology* 2002;**13**:685–92.
- 18 **Agalliu I**, Kriebel D, Quinn MM, *et al.* Prostate cancer incidence in relation to time windows of exposure to metalworking fluids in the auto industry. *Epidemiology* 2005;**16**:664–71.
- 19 **Hastie TJ**, Tibshirani RM. *Generalized additive models*. London: Chapman & Hall, 1990.
- 20 **Ashe HB**, Bergstrom DE. Twenty-six years' experience with dust control in the Vermont granite industry. *Ind Med Surgery* 1964;**33**:73–78.
- 21 **Graham WG**. Re: Attfield M, Costello J. Quantitative exposure-response for silica dust and lung cancer in Vermont granite workers. *Am J Ind Med* 2004;**45**:129–38, *Am J Ind Med* 2004;**46**:89; author reply 90.
- 22 **Stayner L**, Smith R, Schnorr T, *et al.* Re: Lamm *et al.* Determinants of lung cancer risk among cadmium-exposed workers. *Ann Epidemiol* 1993;**3**:114–16.
- 23 **Viren J**, Silvers A. Nonlinearity in the lung cancer dose-response for airborne arsenic: Apparent confounding by year of hire in evaluating lung cancer risks from arsenic exposure in Tacoma smelter workers. *Regul Toxicol Pharm* 1999;**30**:117–29.
- 24 **Gilbert ES**. Some confounding factors in the study of mortality and occupational exposures. *Am J Epidemiol* 1982;**116**:177–88.
- 25 **Hernan MA**, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;**15**:615–25.
- 26 **Eisen EA**, Robins JM. Healthy worker effect in occupational studies. In: El-Shaarawi A, Piegorisch W, eds. *Encyclopedia of environmentrics*. UK: John Wiley & Sons, 2001;987–91.
- 27 **Robins J**. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *J Chronic Dis* 1987;**40**(Suppl 2):139S–61S.
- 28 **Robins JM**. A new approach to causal inference in mortality studies with sustained exposure periods: application to control of the healthy worker survivor effect (erratum appear in *Math Modelling* 1987;**14**:917–21). *Mathematical Modelling* 1987;**7**:1393–512.
- 29 **Robins JM**, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;**11**:550–60.
- 30 **Checkoway H**, Pearce N, Kriebel D. *Research methods in occupational epidemiology. Monographs in epidemiology and biostatistics; v. 34*. Second edition. New York: Oxford University Press, 2004.
- 31 **Asp S**. Confounding by variable smoking habits in different occupational groups. *Scand J Work Environ Health* 1984;**10**:325–26.
- 32 **Kriebel D**, Zeka A, Eisen EA, *et al.* Quantitative evaluation of the effects of uncontrolled confounding by alcohol and tobacco in occupational cancer studies. *Int J Epidemiol* 2004;**33**:1040–5.
- 33 **Steenland K**, Greenland S. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *Am J Epidemiol* 2004;**160**:384–92.
- 34 **Blair A**, Stewart P, Lubin JH, *et al.* Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. *Am J Ind Med* 2007;**50**:199–207.
- 35 **Choi BC**. Definition, sources, magnitude, effect modifiers, and strategies of reduction of the healthy worker effect. *J Occup Med* 1992;**34**:979–88.
- 36 **Steenland K**, Stayner L. The importance of employment status in occupational cohort mortality studies. *Epidemiology* 1991;**2**:418–21.
- 37 **Checkoway H**, Franzblau A. Is silicosis required for silica-associated lung cancer? *Am J Ind Med* 2000;**37**:252–9.
- 38 **Theriault GP**, Peters JM, Johnson WM. Pulmonary function and roentgenographic changes in granite dust exposure. *Arch Environ Health* 1974;**28**:23–7.

Take advantage of BMJ Journals' remarkable catalogue of titles with Related Collections

No busy professional has time to browse through all pertinent journals to find relevant articles, but with Related Collections you no longer have to. Follow the "Related Collections" link from any article and use the "Show Collections from other Journals" to expand your search across all BMJ Journals. Or simply follow the "Browse by topic" link on the home page. By setting up your own collections and receiving email alerts every time an article is added to your chosen area, you can build up your own significant body of knowledge.