



OPEN ACCESS

## ORIGINAL ARTICLE

# Lung-cancer mortality among Vermont granite workers: factors influencing exposure–response evaluation

Pamela M Vacek,<sup>1</sup> Peter W Callas<sup>2</sup>

<sup>1</sup>Medical Biostatistics Unit and Department of Pathology, University of Vermont College of Medicine, Burlington, Vermont, USA

<sup>2</sup>Department of Mathematics and Statistics, University of Vermont, Burlington, Vermont, USA

## Correspondence to

Dr Pamela M Vacek, Medical Biostatistics University of Vermont Hills Building 105 Carrigan Drive Burlington, VT 05405, USA; pvacek@uvm.edu

Received 4 January 2016

Revised 27 September 2016

Accepted 12 October 2016

Published Online First

14 November 2016

## ABSTRACT

**Objective** To understand why 2 studies relating crystalline silica exposure to lung-cancer mortality in Vermont granite workers yielded conflicting results.

**Methods** Data used in the 2 studies were linked to identify discrepancies. Mortality data and employment histories from the earlier study were revised based on data obtained in the later study. SMR were computed and Poisson regressions corresponding to those in the earlier study were performed using the original and revised data. Analyses were repeated with the addition of workers omitted from the earlier study.

**Results** After correction of incomplete mortality and employment information in the original data, the overall SMR for the cohort in the earlier study increased from 1.17 (95% CI 1.03 to 1.36) to 1.39 (95% CI 1.22 to 1.59), and was similar to the SMR of 1.37 observed in the later study (95% CI 1.23 to 1.52). The exposure–response relationship was attenuated, particularly when person-years in all exposure categories were included in the analysis. Inclusion of additional workers had a smaller impact on the SMRs but further attenuated the exposure–response relationship.

**Conclusions** Differing results from the 2 studies are partly attributable to incomplete vital status and work history information used in the earlier study, as well as differences in cohort inclusion criteria. However, differences in length of follow-up and other factors likely play a larger role.

## INTRODUCTION

Studies examining the relationship between occupational exposure to crystalline silica and lung cancer have yielded inconsistent results, even when conducted in the same industry. Two published studies of Vermont granite workers are a notable example of this. An exposure–response analysis conducted by Attfield and Costello,<sup>1</sup> which used data compiled by Graham *et al*,<sup>2</sup> found lung-cancer mortality was increased compared with national death rates (SMR=1.18) and was significantly related to silica exposure. Our group conducted a subsequent study and paradoxically found a greater elevation in lung-cancer mortality (SMR=1.37) but no evidence of a relationship with exposure.<sup>3</sup> Both studies used records from the Vermont Department of Health, Division of Industrial Hygiene (DIH) radiographic surveillance programme, which began obtaining periodic chest X-rays from granite workers in 1937 and continued to do so through

## What this paper adds

- ▶ Quantitative estimates of the relationship between exposure to crystalline silica and lung cancer risk have varied widely, even among studies conducted in the same industry.
- ▶ One study of Vermont granite workers found a modest increase in overall lung-cancer mortality that was related to silica exposure, while a subsequent study found a higher increase that was not related to exposure.
- ▶ This investigation indicated that the conflicting findings were partly attributable to incomplete mortality and work history information used in the earlier study, but that length of follow-up, number of lung cancer deaths, and other factors likely played a larger role.
- ▶ The results demonstrate some of the difficulties in using epidemiological data to estimate exposure–response and determine appropriate exposure limits.

1982.<sup>4</sup> As part of this programme, self-reported work histories were obtained at the time of X-ray and these formed the basis for exposure estimation in both the studies.

There were, however, a number of differences in the methods and data used in the two studies. These are summarised in table 1 and include differences in cohort size and eligibility criteria, length of follow-up, mortality assessment, work history compilation, exposure estimation and statistical analysis. Our study had a larger cohort, partly because it included more recent workers, and also because we did not restrict our cohort to men who participated in the DIH surveillance programme. In addition to using the DIH records, we identified workers from applications for group insurance, which has been available to workers since 1947, pension records for workers employed since 1957, data from a longitudinal study of workers employed between 1979 and 1987<sup>5 6</sup> and data from a study of retired workers.<sup>7</sup> It was previously assumed that participation in the DIH surveillance programme was about 98%. This was based on a paper by Ashe and Bergstrom,<sup>4</sup> who reported that 98.3% of the men who were working on clinic day in 1963 had been X-rayed at least once as part of the DIH surveillance programme. However, our



CrossMark

**To cite:** Vacek PM, Callas PW. *Occup Environ Med* 2017;**74**:211–217.

**Table 1** Methods and data used in two studies of Vermont granite workers

|                               | Attfield and Costello*  | Vacek <i>et al</i>   |
|-------------------------------|---|--|
| Cohort size                   | 5408  | 7052   |
| Eligibility criteria          | Male<br>Employed 1950–1982<br>X-rayed by DIH  | Male<br>Employed 1947–1998   |
| Follow-up                     | Through 1994  | Through 2004   |
| Lung cancer deaths            | 208   | 359  |
| Ascertainment of vital status | Vermont vital records<br>DIH listing of deaths<br>Social Security death index<br>Current employee listings<br>Current pension recipient listings<br>Personal contacts | NDI search<br>SSA vital status search<br>Quebec vital records search<br>Vermont vital records<br>SS death index<br>Obituary and genealogy websites |
| Data sources                  | DIH surveillance programme records  | DIH surveillance programme records<br>Group insurance records<br>Pension records<br>Other studies <sup>4–6</sup>                                   |
| Job-exposure matrix           | Modification of Davis <i>et al</i>  | Verma <i>et al</i>   |
| Dust measurements conversion  | 1 mppcf=0.0075 mg/m <sup>3</sup> silica   | 1 mppcf=0.010 mg/m <sup>3</sup> silica   |
| Exposure–response analysis    | Poisson regression  | Conditional logistic regression  |

\*Based on data collected and compiled by Graham *et al*.<sup>2</sup>

DIH, Division of Industrial Hygiene; mppcf, millions of particles per cubic foot; NDI, National Death Index; SS, Social Security; SSA, Social Security Administration.

data indicated that 18.3% of all men employed in the Vermont granite industry during 1950–1982, the period used to define the cohort in the Attfield and Costello study, did not appear in the DIH records.

The pension and insurance records used in our study were generously made available to us by the workers' union. The records allowed more complete identification of the workforce and provided Social Security numbers for most of the workers. This enabled us to conduct searches of the US National Death Index (NDI) and Social Security Administration (SSA) vital status records. These searches, as well as 10 more years of follow-up, yielded 356 lung cancer deaths for the exposure–response analysis, compared with 201 in the Attfield and Costello analysis. The pension records were also useful for constructing employment histories because they contain information about the company employing the worker and the hours worked for each month since May, 1957. The records also show the years of credit for earlier service in the Vermont granite industry. They thus provided accurate dates of employment for men working after 1 May 1957, and were used to validate or update the self-reported employment information in the DIH records and pulmonary function study of workers employed from 1979 to 1987.<sup>6</sup>

The two studies also used different job-exposure matrices (JEMs) to estimate exposures and different statistical analyses to model exposure–response. Attfield and Costello used a JEM developed by Davis *et al*,<sup>8</sup> while we developed our own JEM.<sup>9</sup> They categorised person-years of follow-up into age, calendar year and cumulative exposure groups, and used Poisson regression to model the observed numbers of lung cancer deaths. We matched cases to controls who were born the same year and survived longer than the case, and performed conditional logistic regression using individual cumulative exposure estimates.

To better understand the differing results obtained in our study and that of Attfield and Costello, we repeated their analysis using the revised mortality and employment information obtained in our study. We also examined how results were impacted by the addition of workers who did not participate in the DIH surveillance programme.

## METHODS

Data from the two studies were linked to identify discrepancies in dates of birth, dates of death, causes of death, hire dates and dates of last employment. Data from our study were also used to identify men who worked in the Vermont granite industry between 1950 and 1982 but were not included in the Attfield and Costello cohort because they did not participate in the DIH surveillance programme. Exposures were estimated using the work histories and methods described by Attfield and Costello.<sup>1</sup> Briefly, the work history data includes codes for 149 different job descriptions and each is accompanied by a location code indicating whether the job was in a quarry or shed. The jobs were classified into 17 shed and 5 quarry categories corresponding to the JEM developed by Davis *et al*.<sup>8</sup> Attfield and Costello modified the JEM by using a conversion factor of 0.0075 to convert dust concentration in millions of particles per cubic foot (mppcf) to mg/m<sup>3</sup> of respirable free silica. They also split the two exposure periods (pre vs post dust controls) in the Davis *et al* JEM into three (pre-1940, 1940–1950 and post-1950) to reflect the fact that dust controls were phased in over time. Exposure concentrations for the 1940–1950 time period were obtained by averaging the pre and post dust control exposure estimates in the Davis *et al* JEM.

The person-year and Poisson regressions analyses described in Attfield and Costello were performed before and after revisions to birth date, date of death, cause of death and employment information. The analyses were then repeated with the addition of workers who were not included in the Attfield and Costello study because they did not have DIH data. The Occupational Cohort Mortality Analysis Program (OCMAP Plus) was used to perform modified life table analyses and to generate the data needed for Poisson regression analyses.<sup>10</sup> As in the analysis by Attfield and Costello, follow-up started 15 years after first employment or in 1950 if this date was earlier and was censored at the end of 1994 if the worker had not died before that time. SMRs were computed based on rates for white males in the US population for the corresponding 5-year age and calendar year groupings. Person-years analyses were also performed with the Life Table Analysis Software (LTAS)<sup>11</sup> used

by Attfield and Costello, to verify that both produced the same results. SAS Proc GENMOD was used to perform the Poisson regression analyses (SAS Institute Inc. SAS/STAT® 9.2 User's Guide, Cary, NC: SAS Institute Inc.; 2008). In externally controlled analyses, population mortality rates were included in the model to adjust for the effects of age and year, while the internally controlled analyses included 5-year age and calendar year groupings, as well as their interaction, in the model.

## RESULTS

### Comparison of mortality data

In the original data from Graham *et al*<sup>2</sup> used by Attfield and Costello<sup>1</sup> there were 5408 persons and 2506 were identified as deceased by the end of 1994, 209 of whom had lung cancer listed as the cause of death. Comparison with the data used in our study indicated that 20 workers had duplicate records and 2 were women. One of the duplicate records was a lung cancer case, so excluding these 22 records 5386 men and 208 lung cancers remained. The mortality data acquired for our study indicated that 174 men who were listed as alive in the data used by Attfield and Costello had actually died by the end of 1994, two men who were listed as deceased were still alive at the end of 1994, and an additional 55 men had unknown vital status, as determined by the SSA vital status search. The additional deaths included 20 lung cancer cases. There were also six lung cancers identified among men with an unknown cause of death in the data used by Attfield and Costello, and one deceased worker who had been erroneously identified as a lung cancer case. Thus, in the corrected data 2671 of the 5386 workers in the cohort were deceased by the end of 1994 and 233 had died of lung cancer.

Among 2484 workers who were classified as being deceased by the end of 1994 in both our data and the Attfield and Costello data, the dates of death agreed within 1-year for all except 13 workers. Five of these larger discrepancies were 10 or more years apart, with one of the deaths occurring 40 years later than indicated in the original data. Information on birth dates generally showed good agreement between the two studies, with only 10 workers having discrepancies of 3 or more years and 3 having discrepancies of 10 or more years. Most of the large discrepancies occurred because two or more workers had the same name, which sometimes resulted in the mortality information being assigned to the wrong person. The Social Security numbers available for our study made it easier to distinguish between workers with the same or similar names.

### Comparison of employment dates

Comparison of year of hire from the pension records used in our study and the DIH data identified 325 discrepancies. Although 147 (46%) of the discrepancies were within 2 years, other differences were as large as 27 years. For all except 18 workers, the hire dates in the pension records were earlier than those in the self-reported DIH data. Employment end dates differed for 2560 (48%) of the cohort members. Nearly half of these discrepancies (1201) occurred because the DIH employment histories ended in 1982, when the surveillance programme was terminated, and thus subsequent employment for these workers was not included in the exposure assessments used in the analyses published by Attfield and Costello. The remaining 1359 discrepancies ranged from -28 to +26 years and for the majority (76%) the end dates were later in the pension data than the DIH data. Nearly half of the discrepancies (47%) were within 2 years, which may reflect inaccurate recall in the self-reported work histories.

On the basis of the DIH data, 117 of the 5386 men remaining in the cohort, after elimination of the data of women and duplicates, had stopped working in the Vermont granite industry before 1950 and thus should not have been included in the cohort. However, the pension data showed that only 42 of the 117 workers had left the industry before 1950, with end dates ranging from 1940 to 1949. None of these workers were excluded from the analysis published by Attfield and Costello, and we therefore did not exclude them from any of the analyses presented in this paper. We did, however, examine how their exclusion impacted the results and found that it had little effect.

### Identification of additional workers

We identified 1205 male workers who were employed in the Vermont granite industry from 1950 to 1982 but did not have an X-ray as part of the DIH surveillance programme, yielding an expanded cohort of 6591 workers. Birth years for the additional workers ranged from 1882 to 1966 and year of hire ranged from 1902 to 1982. Their mean year of birth was 1939, compared with 1921 for the original cohort and they consequently began work later, with a mean hire year of 1967 compared with 1947 for the original cohort. The majority (89%) were hired in 1950 or later and the types of jobs listed on their work histories were similar to the original cohort. Twenty-three of the additional workers had died of lung cancer by the end of 1994.

### Reanalysis of exposure-response

The results of the person-years analysis without an exposure lag are shown in table 2. Some of the results based on the original data (table 2) differ slightly from those presented by Attfield and Costello, although the numbers of observed and expected lung cancers in each exposure category are the same as in their results. The differences are indicated in bold type and include changes to four SMRs and the total number of expected lung cancers, which were inconsistent with the observed and expected lung cancers in table IV of Attfield and Costello.<sup>1</sup> We also believe that the cut-off point for the highest exposure category should be 5.0 mg/m<sup>3</sup>-years rather than 6.0 mg/m<sup>3</sup>-years, as shown in their table, because we could not replicate the results using a higher cut-off point. They used the cut-off point 5.0 mg/m<sup>3</sup>-years in table III of their paper and we did replicate the cumulative exposure distribution presented there. Attfield and Costello did not indicate what age by year distribution of person-years they used to compute the standardised rate ratios (SRRs) and we were unable to replicate their results using either the distribution in the lowest category or the overall distribution for the cohort. The age by year distribution is highly confounded with exposure category so the choice of reference distribution had a large impact on the SRRs. We therefore have not presented these results.

After revision of cohort, mortality data and employment information (table 2) SMRs were larger in most exposure categories, with the biggest difference being observed in the 1.0 mg/m<sup>3</sup>-years category (1.80 vs 1.27). The overall SMR (1.39) was comparable to that observed in our study (1.37).<sup>3</sup> Inclusion of additional workers further increased the SMR in the lowest exposure category but had little or no effect on the SMRs for the higher exposure categories (table 2).

SMRs from person-years analyses that included a 15 year exposure lag are shown in table 3. Inclusion of a lag increased the SMRs for some exposure categories and reduced it for others, with a general attenuation of differences. However, unlike the analysis based on the original data, use of an

**Table 2** SMRs for lung cancer by exposure category

|   | Cumulative exposure mg/m <sup>3</sup> -years |             |             |             |         |             |        |         |
|---|--|-------------|-------------|-------------|---------|-------------|--------|---------|
|   | 0—   | 0.25—       | 0.50—       | 1.00—       | 1.50—   | 2.00—       | 3.00—  | 5.00—   |
| (A) Original data*                                |  |             |             |             |         |             |        | All     |
| Person-years                                      | 29756.6                                      | 10148.0     | 15001.5     | 12910.5     | 9098.8  | 9726.3      | 8289.8 | 13171.4 |
| Observed deaths                                   | 29   | 15          | 28          | 22          | 22      | 26          | 29     | 30      |
| Expected deaths                                   | 37.8   | 16.8        | 20.8        | 17.3        | 16.5    | 17.6        | 17.1   | 25.8    |
| SMR   | 0.77   | <b>0.89</b> | <b>1.35</b> | <b>1.27</b> | 1.33    | <b>1.48</b> | 1.70   | 1.16    |
| (B) Revised cohort, mortality and employment data |  |             |             |             |         |             |        | All     |
| Person-years                                      | 24309.2                                      | 11271.0     | 14509.1     | 12996.1     | 9799.2  | 10365.5     | 9034.8 | 13513.7 |
| Observed deaths                                   | 28   | 22          | 27          | 29          | 22      | 27          | 40     | 29      |
| Expected deaths                                   | 27.6   | 18.5        | 19.1        | 16.1        | 16.0    | 18.8        | 18.4   | 26.1    |
| SMR   | 1.01   | 1.19        | 1.41        | 1.80        | 1.37    | 1.43        | 2.17   | 1.11    |
| (C) Inclusion of additional workers               |  |             |             |             |         |             |        | All     |
| Person-years                                      | 31539.7                                      | 12626.6     | 15086.1     | 13440.6     | 10149.7 | 10564.3     | 9034.8 | 13513.7 |
| Observed deaths                                   | 43   | 24          | 27          | 30          | 24      | 27          | 40     | 29      |
| Expected deaths                                   | 36.4   | 19.9        | 20.2        | 16.9        | 16.7    | 19.4        | 18.4   | 26.1    |
| SMR   | 1.18   | 1.20        | 1.33        | 1.77        | 1.44    | 1.39        | 2.17   | 1.11    |

\*Numbers in bold differ from those in table IV of Attfield and Costello but are in agreement with the numbers of observed and expected deaths presented in their table.

exposure lag with the revised data increased the SMR in the lowest exposure category.

The regression results corresponding to the analyses presented in table VII of Attfield and Costello<sup>1</sup> are shown in table 4. As in their study, regressions were performed on both untransformed and log-transformed cumulative exposure, using either internal or external adjustment for age and calendar year, and were repeated after the exclusion of the highest exposure category. The results for the internally controlled analyses based on the original data are slightly different from those of Attfield and Costello because they appear to have included the 5-year age group and decade in the regression model, while we included 5-year age and calendar year groups, together with their interaction, to adjust for the individual age-year strata. After revisions to the data, no significant associations between exposure and lung-cancer mortality were observed when all exposure

groups were included in the analysis. Inclusion of additional workers yielded similar, non-significant results. As in the findings reported by Attfield and Costello, when the highest exposure category was excluded from the analyses, significant exposure-response relationships were observed in all analyses. However, the regression coefficients were somewhat lower than in the results presented by Attfield and Costello, especially when additional workers were included in the analysis.

## DISCUSSION

The analyses performed in this study indicate that differences in mortality ascertainment, employment data and eligibility criteria contributed to differences in the results reported by Attfield and Costello and those from our study but did not entirely explain them. The addition of deaths that were not captured in the original data increased the SMRs in all exposure categories,

**Table 3** SMRs for lung cancer by cumulative exposure category—15 year exposure lag

|   | Cumulative exposure mg/m <sup>3</sup> -years |         |         |        |        |        |        |         |
|---|--|---------|---------|--------|--------|--------|--------|---------|
|   | 0—   | 0.25—   | 0.50—   | 1.00—  | 1.50—  | 2.00—  | 3.00—  | 5.00—   |
| (A) Original data                                 |  |         |         |        |        |        |        | All     |
| Person-years                                      | 41284.7                                      | 11909.2 | 14747.6 | 9667.8 | 6323.5 | 6875.3 | 6394.1 | 10900.9 |
| Observed deaths                                   | 35   | 18      | 32      | 22     | 17     | 27     | 23     | 27      |
| Expected deaths                                   | 45.0   | 17.9    | 22.5    | 18.0   | 13.8   | 14.4   | 14.0   | 24.2    |
| SMR   | 0.78   | 1.01    | 1.42    | 1.22   | 1.23   | 1.88   | 1.64   | 1.12    |
| (B) Revised cohort, mortality and employment data |  |         |         |        |        |        |        | All     |
| Person-years                                      | 36167.0                                      | 14251.8 | 14905.0 | 9317.1 | 6268.0 | 7181.4 | 6607.6 | 11100.0 |
| Observed deaths                                   | 39   | 22      | 31      | 28     | 20     | 32     | 26     | 26      |
| Expected deaths                                   | 35.2   | 20.3    | 21.4    | 17.0   | 13.1   | 15.0   | 14.3   | 24.3    |
| SMR   | 1.11   | 1.08    | 1.45    | 1.65   | 1.52   | 2.13   | 1.81   | 1.07    |
| (C) Inclusion of additional workers               |  |         |         |        |        |        |        | All     |
| Person-years                                      | 44488.9                                      | 14857.0 | 15395.0 | 9744.8 | 6475.7 | 7289.6 | 6607.6 | 11100.0 |
| Observed deaths                                   | 54   | 24      | 31      | 30     | 21     | 32     | 26     | 26      |
| Expected deaths                                   | 44.5   | 21.5    | 22.4    | 17.9   | 13.6   | 15.4   | 14.3   | 24.3    |
| SMR   | 1.21   | 1.11    | 1.38    | 1.67   | 1.54   | 2.08   | 1.81   | 1.07    |



**Table 4** Exposure–response coefficients from Poisson regression of lung-cancer mortality rates

|   | Internal adjustment |         |               |         | External adjustment |         |               |         |
|---|---------------------|---------|---------------|---------|---------------------|---------|---------------|---------|
|   | Cumulative*         |         | ln Cumulative |         | Cumulative          |         | ln Cumulative |         |
|   | Coefficient†        | p Value | Coefficient   | p Value | Coefficient         | p Value | Coefficient   | p Value |
| <i>Exposure unlagged</i>                      |                     |         |               |         |                     |         |               |         |
| Including all exposure groups                 |                     |         |               |         |                     |         |               |         |
| Original data                                 | 0.013               | 0.543   | 0.144         | 0.006   | 0.017               | 0.371   | 0.114         | 0.008   |
| Revised cohort, mortality and employment data | −0.001              | 0.951   | 0.089         | 0.076   | −0.006              | 0.771   | 0.063         | 0.148   |
| Inclusion of additional workers               | −0.006              | 0.761   | 0.056         | 0.223   | −0.007              | 0.600   | 0.043         | 0.126   |
| Excluding last exposure group                 |                     |         |               |         |                     |         |               |         |
| Original data                                 | 0.204               | 0.001   | 0.230         | 0.001   | 0.180               | 0.002   | 0.180         | 0.001   |
| Revised cohort, mortality and employment data | 0.172               | 0.003   | 0.175         | 0.004   | 0.161               | 0.003   | 0.167         | 0.004   |
| Inclusion of additional workers               | 0.149               | 0.007   | 0.126         | 0.021   | 0.142               | <0.001  | 0.124         | 0.007   |
| <i>Exposure lagged 15 years</i>               |                     |         |               |         |                     |         |               |         |
| Including last exposure group                 |                     |         |               |         |                     |         |               |         |
| Original data                                 | 0.008               | 0.719   | 0.106         | 0.021   | 0.017               | 0.402   | 0.091         | 0.007   |
| Revised cohort, mortality and employment data | −0.008              | 0.726   | 0.077         | 0.113   | −0.010              | 0.610   | 0.056         | 0.180   |
| Inclusion of additional workers               | −0.013              | 0.550   | 0.049         | 0.271   | −0.011              | 0.566   | 0.042         | 0.282   |
| Excluding last exposure group                 |                     |         |               |         |                     |         |               |         |
| Original data                                 | 0.204               | 0.002   | 0.170         | 0.001   | 0.190               | 0.001   | 0.144         | 0.001   |
| Revised cohort, mortality and employment data | 0.159               | 0.009   | 0.160         | 0.007   | 0.153               | 0.005   | 0.159         | 0.004   |
| Inclusion of additional workers               | 0.136               | 0.018   | 0.121         | 0.025   | 0.139               | 0.009   | 0.127         | 0.012   |

\*Cumulative exposure (mg/m<sup>3</sup>-years) included in the model as the mean of each exposure category or the natural logarithm (ln cumulative) of the mean.

†Regression coefficient for the increase in the natural logarithm of relative risk per unit increase in cumulative exposure or its logarithm.

yielding an overall SMR of 1.39. This is considerably higher than the SMR of 1.17 reported by Attfield and Costello<sup>1</sup> and similar to the SMR of 1.37 obtained in our mortality study.<sup>3</sup> After correction of both mortality and employment data, the SMRs based on the revised data did not show as consistent a trend across exposure categories as the analyses reported by Attfield and Costello.<sup>1</sup> This was particularly the case for the analysis based on unlagged exposures because in the original data the last year of employment was recorded as 1982 for the 1201 men still working at that time. Follow-up continued through 1994, so unlagged exposures for workers employed after 1982 were underestimated.

The addition of workers who were not included in the Attfield and Costello cohort because they did not have an X-ray as part of the voluntary DIH surveillance programme increased the SMR in the lowest exposure category from 1.01 to 1.18 in the analysis using unlagged exposures and from 1.11 to 1.21 when exposures were lagged 15 years. There was little effect on the SMRs in the other exposure categories because 89% of the additional workers were hired in 1950 or later and hence a large portion of their person-years of follow-up were in the lowest exposure category. It is unclear why they had increased lung-cancer mortality, but perhaps smokers with low silica exposure were less likely to participate in the DIH surveillance programme than non-smokers.

After revisions were made to the original mortality and employment data, the Poisson regression estimates for the association between cumulative exposure and the relative risk of lung-cancer mortality were generally attenuated and the regression coefficients were further reduced when additional workers were included. This is largely attributable to the increased SMR in the lowest exposure category. An exception was the analysis based on lagged exposures, a logarithmic transformation of cumulative exposure, no additional workers and exclusion of the highest exposure categories (table 4). In this case, the higher mortality in the lowest categories was offset by increased

mortality in the two categories preceding the highest one. None of the regression results were statistically significant when all exposure categories were included in the analysis but most were significant when the highest category was excluded, consistent with the results reported by Attfield and Costello.<sup>1</sup> In our mortality study we did not exclude men who attain high cumulative exposures from the nested case–control analysis, but we did fit a number of non-linear exposure–response relationships, including polynomial and spline functions, and found no significant associations.<sup>3</sup> We also found no association when we excluded men who were born before 1920 from the analysis,<sup>3</sup> which eliminated the exposures in Attfield and Costello's highest category.

Post hoc exclusion of data is always problematic because it alters the interpretation of the significance levels obtained from subsequent analyses. Nevertheless, a reduction in the effects of exposure at high levels has been observed in other epidemiological studies and can be due to a number of causes.<sup>12</sup> In the Attfield and Costello study,<sup>1</sup> one rationale for excluding the highest exposure category was that the exposure estimates were thought to be the poorest, although most of the person-years of exposure in the penultimate exposure category were based on the same exposure estimates. Another explanation for the lower mortality observed in the highest exposure category is that it includes healthier workers, who were able to work long enough to attain high cumulative exposure (a healthy worker survivor effect). To explore this possibility we split the cohort into men who started work before 1950 (prevalent hires) and those who started in or after 1950 (incident hires) and repeated the SMR and Poisson regression analyses using the revised data with an exposure lag of 15 years, both with and without the additional workers. The results for the prevalent hires were similar to those shown in table 4 for the full cohort, with significant associations observed only when the highest exposure category was excluded. For the incident hires, most person-years of follow-up were in the lowest exposure category and there were no

exposures in the two highest two categories. Significance levels for the regression results ranged from 0.151 to 0.510, although the coefficients based on a log transformation of cumulative exposure did not differ widely from those for the prevalent hires when the highest exposure category was excluded (0.179 vs 0.156 without the additional workers).

Applebaum *et al*<sup>13</sup> performed similar analyses on the data from the Attfield and Costello study, using Cox regression and categorising cumulative exposure based on the quartiles of the cumulative exposure distribution for the cases from the full cohort. Their highest quartile included exposures over 2.9 mg/m<sup>3</sup>-years, essentially combining Attfield and Costello's two highest categories, and they found a significant trend for both the prevalent and incident hires. Applebaum *et al* repeated their analyses using cut points based on the quartiles for cases in the prevalent hires and tertiles for the incident hires. They found a significant trend across exposure categories for the incident hires, but not for the prevalent hires. We performed Poisson regression on the revised data without additional workers using these same cut points and found no significant trend for either the prevalent ( $p=0.375$ ) or incident hires ( $p=0.175$ ). The differences between our results and those of Applebaum *et al* may be due to the missed deaths in their data, particularly in the lower exposure categories, and the fact that they controlled for age at hire. In addition, the subcohorts in their analyses were not identical to ours because corrections to hire dates switched some workers from the incident to prevalent subcohort.

The exposure–response results for prevalent and incident hires based on the revised data do not provide much evidence for a healthy worker survivor effect, but they do highlight the problem of combining two distinct subcohorts with widely differing exposures. All of the cumulative exposures over 3.0 mg/m<sup>3</sup>-years occurred in men employed prior to 1950, while most of the cumulative exposures under 0.5 mg/m<sup>3</sup>-years occurred in men hired after that date. The regression results, which are highly influenced by the difference between the lowest and highest exposure categories, are thus based on comparing cohorts of men whose person-years of follow-up differ widely in age, calendar year and exposure. This confounding is particularly problematic for assessing the effects of exposure on lung cancer, because a nationwide increase in cigarette smoking occurred at the same time silica exposures in the industry were reduced. Neither our mortality study nor that of Attfield and Costello had smoking information on the workers.

There were a number of differences between the two studies that were not explored in this investigation. In addition to the 1205 men who worked between 1950 and 1982 but did not participate in the DIH surveillance programme, our cohort included 461 men who only worked between either 1947 and 1949 or 1983 and 1998. However, there were no lung cancer deaths among these workers, so they had a negligible effect on the case–control analyses used in our study. Our study also had longer follow-up and this may have contributed substantially to the difference between the exposure–response results from the two studies. During 10 more years follow-up, 115 lung cancer deaths occurred in the cohort that included the original and additional workers employed from 1950 to 1982. Many of these were among men with lower exposure and thus are likely to have had a large impact on the exposure–response relationship.

The exposure estimates for the two studies also differed. We found many gaps in employment that were not captured by the DIH work histories. Although we corrected employment start and stop dates, we did not correct for gaps in employment when revising the data for the reanalysis presented in this paper

because it was not feasible to reconstruct work histories for the entire cohort. In addition, we developed our own JEM based on 5204 exposure measurements made between 1924 and 2004 and used a factor of 0.010 to convert dust concentrations (mppcf) to mg/m<sup>3</sup> respirable silica.<sup>9</sup> Nevertheless, the job categories and exposure estimates were generally similar to those of Attfield and Costello, which were based on the Davis *et al*<sup>8</sup> JEM and a conversion factor of 0.0075 mg/m<sup>3</sup> per mppcf. Notable exceptions were the pre-1940 estimates for channel bar operators and plug drillers. Attfield and Costello used an exposure estimate of 1.07 mg/m<sup>3</sup> for channel bar operators and all drillers except plug drillers, which were assigned an exposure of 0.65 mg/m<sup>3</sup>. We used an estimate of 0.15 mg/m<sup>3</sup> for channel bar operators, based on measurements made during the 1930s at two quarries using wet processes for this activity, and an estimate of 1.07 mg/m<sup>3</sup> for all drillers, including Leyner drillers and plug drill operators. As part of our mortality study, we had performed a series of sensitivity analyses using alternative exposures for these jobs. When we used the estimate of 1.07 mg/m<sup>3</sup> for pre-1940 channel bar operators, it had virtually no effect on our exposure–response results.<sup>3</sup> This was not surprising because most of the person-years of employment for channel bar operators occurred after 1950, when our exposure estimate was identical to that used by Attfield and Costello, and only one channel bar operator with substantial employment prior to 1940 died of lung cancer. In our original sensitivity analysis, plug drillers were not assigned the lower exposure used by Attfield and Costello, so we recently repeated this analysis using their definitions and pre-1940 estimates for plug drillers, in addition to those for channel bar workers, Leyner drillers and other drillers. Again, we found no evidence of an exposure–response relationship for lung cancer ( $OR=0.98$ ,  $p=0.37$ ). As we reported previously, the strongest exposure–response relationship for silicosis was obtained when a pre-1940 exposure of 0.15 mg/m<sup>3</sup> was used for all these workers,<sup>3</sup> suggesting that Attfield and Costello's JEM may overestimate exposure for channel bar operators and all drillers, except perhaps plug drillers, while ours may overestimate exposures for all drillers. Our use of a differing conversion factor when constructing the JEM would alter the slope of the exposure–response relationship by a proportional amount but would not reduce the significance of the association.

This investigation demonstrates the difficulty in using historical cohort data for exposure–response analysis, especially when the data covers a long time period. The confounding of age, calendar year and exposure observed in the Vermont granite data is likely to occur in other occupational studies that span time periods during which exposure controls were introduced and/or improved, complicating the interpretation of results.

**Correction notice** This paper has been updated since it first published online. Instances of 'mg/m<sup>3</sup>/years' have been corrected to 'mg/m<sup>3</sup>-years'.

**Contributors** PMV contributed to the study design, analysis and interpretation of data, and drafted the manuscript. PWC contributed to the analysis and interpretation of data, and critical revision of the manuscript. Both authors approved the final draft and are accountable for all aspects of the work.

**Funding** This study was partially supported by a gift (R0073) to the University of Vermont from the Research Foundation for Health and Environmental Effects (now the Foundation for Chemistry Research and Initiatives), a non-profit tax-exempt organisation established by the American Chemistry Council (ACC).

**Disclaimer** The sponsor had no involvement in the study design; collection, analysis and interpretation of data; writing of the report; or decision to submit the paper for publication.

**Ethics approval** University of Vermont Research Protections Office, Committees on Human Research.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

## REFERENCES

- 1 Attfeld MD, Costello J. Quantitative exposure-response for silica dust and lung cancer in Vermont granite workers. *Am J Ind Med* 2004;45:129–38.
- 2 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.
- 3 Vacek PM, Verma DK, Graham WG, et al. Mortality in Vermont granite workers and its association with silica exposure. *Occup Environ Med* 2011;68:312–18.
- 4 Ashe HB, Bergstrom DE. Twenty-six years' experience with dust control in the Vermont granite industry. *Ind Med Surg* 1964;33:973–8.
- 5 Graham WGB, Ashikaga T, Hemenway D, et al. Radiographic abnormalities in Vermont granite workers exposed to low levels of granite dust. *Chest* 1991;100:1507–14.
- 6 Graham WGB, Weaver S, Ashikaga T, et al. Longitudinal pulmonary function losses in Vermont granite workers: a reevaluation. *Chest* 1994;106:125–30.
- 7 Graham WGB, Vacek PM, Morgan WKC, et al. Radiographic abnormalities in long-tenure Vermont granite workers and the permissible exposure limit for crystalline silica. *J Occup Environ Med* 2001;43:412–17.
- 8 Davis LK, Wegman DH, Monson RR, et al. Mortality experience of Vermont granite workers. *Am J Ind Med* 1983;4:705–23.
- 9 Verma DK, Vacek PM, des Tombe K, et al. Silica exposure assessment in a mortality study of Vermont granite workers. *J Occup Environ Hygiene* 2011; 8:71–9.
- 10 Marsh GM, Youk AO, Stone RA, et al. OCMAP-PLUS: a program for the comprehensive analysis of occupational cohort data. *J Occup Environ Med* 1998;40:351–62.
- 11 Waxweiler RJ, Beaumont JJ, Henry JA, et al. A modified life-table analysis system for cohort studies. *J Occup Med* 1983;25:115–24.
- 12 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.
- 13 Applebaum KM, Malloy EJ, Eisen EA. Reducing healthy worker survivor bias by restricting date of hire in a cohort study of Vermont granite workers. *Occup Environ Med* 2007;64:681–7.