



Contents lists available at SciVerse ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Sparse-data bias accompanying overly fine stratification in an analysis of beryllium exposure and lung cancer risk

Kenneth J. Rothman DrPH*, Paul L. Mosquin PhD

RTI Health Solutions, Research Triangle Institute, 200 Park Offices Drive, Research Triangle Park, NC

ARTICLE INFO

Article history:

Received 24 May 2012

Accepted 14 November 2012

Keywords:

Epidemiologic methods

Stratified analysis

Sparse-data bias

Beryllium

Lung cancer

ABSTRACT

Purpose: Beryllium's classification as a carcinogen is based on limited human data that show inconsistent associations with lung cancer. Therefore, a thorough examination of those data is warranted. We reanalyzed data from the largest study of occupational beryllium exposure, conducted by the National Institute of Occupational Safety and Health (NIOSH).

Methods: Data had been analyzed using stratification and standardization. We reviewed the strata in the original analysis, and reanalyzed using fewer strata. We also fit a Poisson regression, and analyzed simulated datasets that generated lung cancer cases randomly without regard to exposure.

Results: The strongest association reported in the NIOSH study, a standardized rate ratio for death from lung cancer of 3.68 for the highest versus lowest category of time since first employment, is affected by sparse-data bias, stemming from stratifying 545 lung cancer cases and their associated person-time into 1792 categories. For time since first employment, the measure of beryllium exposure with the strongest reported association with lung cancer, there were no strata without zeroes in at least one of the two contrasting exposure categories. Reanalysis using fewer strata or with regression models gave substantially smaller effect estimates. Simulations confirmed that the original stratified analysis was upwardly biased. Other metrics used in the NIOSH study found weaker associations and were less affected by sparse-data bias.

Conclusions: The strongest association reported in the NIOSH study seems to be biased as a result of non-overlap of data across the numerous strata. Simulation results indicate that most of the effect reported in the NIOSH paper for time since first employment is attributable to sparse-data bias.

© 2013 Elsevier Inc. All rights reserved.

Introduction

Previously we demonstrated [1] through simulation studies and control of age confounding using stratification that the nested case-control study of beryllium and fatal lung cancer by Sanderson et al [2] was affected by strong confounding by year of birth for lagged measures of average daily exposure. Others have also commented on the age confounding in that study [3,4]. Recently, Schubauer-Berigan et al [5] published an update to the National Institute of Occupational Safety and Health (NIOSH) source cohort study within which the case-control study of Sanderson et al was nested; their update summarizes most of the available data that exist on occupational exposure to beryllium and lung cancer risk. The evidence for a relation between beryllium exposure and lung cancer from their study is mixed. They used several metrics to measure beryllium exposure, with varying lag times, and both internal and

external comparisons. The strongest association they reported was for workers with 35 or more years since first employment in the beryllium industry, compared with workers with fewer than 15 years since first employment, for which the standardized rate ratio (SRR) was 3.68. For cumulative exposure, however, the SRR comparing highest quartile with the lowest was only 1.12, based on a 10-year lag, although this value increased to 1.97 after excluding short-term workers. Other measures, such as employment duration (10-year lag) and maximum exposure (unlagged), were also reported, but these associations were smaller.

In this paper, we show that the reported analysis was affected by sparse-data bias, which accounts for most of the reported association between time from first employment in a beryllium plant and lung cancer in the NIOSH cohort study. The NIOSH study included 9199 workers followed from 1940 through 2005 for fatal lung cancer and other endpoints. It encompassed more than 350,000 person-years of follow-up, during which 545 cases of lung cancer were identified. With all these data, it may seem odd that there would be a problem with sparse data. The problem arises because the data were stratified into hundreds of cells, and inferences were drawn from datasets that were populated mostly with zero cell counts.

* Corresponding author. RTI Health Solutions, 200 Park Offices Drive, Research Triangle Park, NC 27709.

E-mail address: KRothman@rti.org (K.J. Rothman).

All ratio measures based on counts, such as those reported by Schubauer-Berigan et al [5], are positively biased on the arithmetic scale because errors that exaggerate the ratio are larger than errors that underestimate it. For a single table, a ratio measure that has a nonzero probability of having a zero denominator will consequently have infinite bias, because the mean estimate will be infinity. (For example, consider the ratio of heads to tails in 10 tosses of a fair coin; although the expectation of the proportion heads is 50%, the expectation for the ratio of heads to tails is infinity, because the outcome of $10/0 = \text{infinity}$ is averaged with other outcomes to get the expected value. Even if the outcome of 10 heads were disallowed by recoding it to 9, the ratio of heads to tails would have an expected value above 1.) With stratified analysis, any unconfounded summary measure is essentially a weighted average of stratum-specific estimates, and is subject to the same problem, which can be exaggerated if the numbers within strata are small. Greenland [6] has suggested that such small-sample bias may be more prevalent than commonly realized. Various solutions may be employed to correct for sparse-data problems in stratified data. It may be possible to collapse neighboring strata without introducing substantial residual confounding. A regression model can be employed that avoids stratifying a continuous variable such as age. In addition, various corrections can be applied to mitigate the bias; two possibilities are the use of the Firth correction [7] and the use of data augmentation to implement Bayesian shrinkage for sparse data [8,9].

We examined the results of Schubauer-Berigan et al in several ways. After replicating their results, we inspected their stratified data, a step that reveals the sparse-data problem. Because the sparse-data problem arises from a combination of fine stratification of the data coupled with non-overlapping exposure distributions, and the purpose of the stratification is to control confounding, we then reanalyzed their data to explore the amount of confounding as well as the magnitude of the sparse-data bias. Finally, we conducted simulations using the actual cohort experience with respect to beryllium exposure, but randomly simulating lung cancer deaths, which enabled us to see the extent to which the finely stratified analysis biased the results.

Methods

NIOSH kindly supplied a copy of the dataset used for this analysis. To verify the data, we first attempted to replicate the results reported in the NIOSH paper. In the NIOSH paper, two analytic approaches were used, both based on stratification to control confounding. One involved external comparison with U.S. mortality data, calculating standardized mortality ratios by exposure level for the cohort. The other was an internal comparison across approximate exposure quartiles in the data, using standardization to summarize the results across strata. The standard used to weight the stratum-specific results was the distribution of person-time in the entire cohort across categories of the stratification variables. There were three stratification variables used: Age, calendar year, and race. Both age and calendar year were categorized into 5-year intervals. For age, there were 16 categories ranging from a low of 10 to 14 years, which had very little person-time, to a high of 85 or older. For calendar year, there were 14 categories, starting with 1940 to 1944 and going to 2005 to 2009. There were two categories of race. The data were further divided by exposure level into approximate quartiles. Several exposure metrics were used; these included employment duration, time since first employment, cumulative beryllium exposure, and maximum beryllium exposure. Most of our analyses focused on time since first employment, the measure that had the largest SRR for lung cancer death (3.68) reported by Schubauer-Berigan et al. Following the approach of Schubauer-Berigan et al., we classified person-time into four

approximately equally sized categories of time (in years) since first exposure, which were bounded as follows from lowest to highest: [0,15) [15,25) [25,35) [35,80). All analyses conducted by NIOSH used publically available cohort analysis software, the Life Table Analysis System (LTAS.NET) [10–12]. We used LTAS.NET for verification but also wrote our own software as a check on LTAS.NET. LTAS.NET uses standard stratification methods to control confounding, coupled with standardization (“direct standardization”) to summarize effects across strata.

After verifying the integrity of the data and the calculations reported by NIOSH, we inspected the strata to assess the distribution of information across exposure levels and strata. We tried alternative stratification schemes to deal with strong confounding, applying the same statistical methods used in LTAS.NET. We used Mantel-Haenszel methods as an alternative to standardization in some calculations. We also fit a Poisson regression model as an alternative to stratification to control confounding without the sparse-data problems inherent in the stratified analysis. In this model, we included terms for age, age-squared, age-cubed, year, year-squared, year-cubed, and race. Regression modeling can also be affected by bias from sparse data, however, so, in alternative analyses, we fit the Poisson regression using the Firth [7] correction, and we used Greenland’s [8,9] approach of Bayesian shrinkage based on data augmentation. For the data augmentation, for each coefficient we used a weak prior that added two pseudo-records, each with one case, and added an indicator for each pair, corresponding to a prior that offers 95% certainty that the rate ratio (RR) is between 0.026 and 39 [9].

In addition, we conducted a series of simulations of the lung cancer findings, by taking the cohort experience and simulating the occurrence of lung cancer deaths. We obtained cause-, calendar year-, race-, and age-specific population mortality rates used with LTAS.NET. For each cohort member, date and cause of mortality was determined randomly by applying the mortality rates to the corresponding amounts of person-time. Simulations using time since first employment and duration of employment were based on all 9199 cohort members; simulations using cumulative exposure and maximum exposure were based on the subcohort of 5436 workers employed at three plants for which linkage with work history data was possible. The simulation methods are described in more detail in the Appendix. The simulation process guaranteed no association between beryllium exposure and death from lung cancer, so that any departure from a null result in the data analysis reflects bias in the methods applied or the estimator used.

In attempting to verify the NIOSH results, we discovered a small problem in the way that follow-up had been defined in the NIOSH study, a problem that led to a “time-loop” [13] and the exclusion of immortal person-time, inflating the estimated rates. The NIOSH protocol considered workers lost to follow-up if they left employment alive and were not ascertained as a death in subsequent follow-up. Thus, the occurrence of a death determined whether the person-time of these retired workers would be included in the study, introducing a selection bias that inflated the mortality rates. This bias would only affect the rate ratio for beryllium exposure and lung cancer mortality if retirement time is related to exposure. That is the case, however, for time since first employment, because the excluded person-time is concentrated among those with the longest time since first employment, inflating the mortality rate most for those in the highest category of time since first employment. This is a time-loop because whether a worker was actually lost to follow-up at the time of retirement from work depended on a future event, whether the worker was ascertained to have died in the Social Security Administration database or the National Death Index (NDI). Fortunately, because follow-up was lengthy and the NDI is nearly complete, only 123 workers were affected by this

aspect of the protocol, making little difference to the analyses. We ignored this issue in trying to replicate the NIOSH results, but in other analyses we addressed it by including in the cohort all follow-up time for retiring workers until they reached age 105, an arbitrary value represent the upper limit of longevity for this very old group of retired survivors.

Results

We were able to verify case counts, person-time totals, standardized mortality ratios, and SRRs by exposure level for all of the published NIOSH results. We then examined the fully stratified data. With 16 age categories, 14 time categories, 2 race categories, and 4 exposure levels for each exposure metric, there were a total of 1792 exposure- and confounder-specific rate categories, over which 545 lung cancer deaths and 352,518 person-years were distributed. Thus, there was an average of only 0.3 cases per exposure cell in the data.

We focused attention on the comparison of the highest versus lowest category of time since first employment. This contrast had the largest SRR for lung cancer death (3.68) reported by Schubauer-Berigan et al. Of 448 strata of age \times calendar year \times race, only 43 strata had nonzero person-time for both the lowest and highest exposure categories. Of the total of 204,579 person-years that were tallied for these two exposure categories, only 32,557, or 16%, fell in strata that had nonzero person-time for both of these two levels of exposure. Of the 448 strata, 370 had no cases in either of these two exposure categories, and of the remaining 78 strata, all had cases in only one of these two exposure categories, with no stratum out of 448 having a nonzero case count in both the lowest and highest quartile of exposure. Thus, every stratum-specific ratio of rates comparing highest with lowest quartile of time since first employment, out of the 78 from which such a ratio could be calculated, gave an estimate of either zero or infinity for the rate ratio. This inspection of the stratified data indicates that sparse data could be a major concern.

The original paper did not report crude results, but we report them here for time since first employment (Table 1). The crude rate ratio estimates are considerably greater than the SRRs reported by the NIOSH investigators, indicating that there was strong confounding by one or more of the stratification variables. For time since first employment, we evaluated the confounding present for age, calendar year, and race in separate analyses, taking each factor into account singly in a stratified analysis and standardizing to the distribution of all person-years within the cohort.

Table 2 shows the results stratifying by age into seven broader categories, rather than the 5-year categories used in the original analysis. Ordinarily, seven strata of a continuous variable would be sufficient to control most of the confounding [14]. Two things are notable from these age-stratified data: (1) There is strong confounding by age, with a SRR of 2.65 and a Mantel-Haenszel [15] estimator of 2.73 for the highest category versus the lowest, compared with the crude RR of 21.5 (Table 1), and (2) even with

stratification into as few as 7 strata, rather than the 448 used in the original study, the highest and lowest category of exposure still have poorly overlapping distributions (Fig. 1). For the lowest category of exposure, nearly half of the total person-time occurs in people younger than 35 years, and only 2% occurs in people 65 or older. For the highest category of exposure, zero person-time occurs in those younger than 35, and 62% occurs in people 65 or older. Thus, the reference category for exposure notably has little person-time in the age categories where most of the person-time for high levels of exposure fell. We conducted parallel analyses stratifying for calendar time and race, but neither of these demonstrated the strong confounding in these data seen for age.

Calendar year and race were confounding in the same direction as age, so it is surprising that the effect estimates for time since first employment in the original publication (3.68 for the highest vs. lowest level of exposure), controlling for all three factors, is considerably greater than the value of 2.65 we found after stratifying only by seven age categories. One would expect that with additional control of the other confounders that confound in the same direction, the effect estimate should be lower than 2.65. One possible explanation for it increasing rather than decreasing as the other confounders are also controlled is bias from sparse data as a result of the extreme stratification. We used three approaches to investigate this possibility: (1) We re-analyzed the data using a Poisson regression model to control simultaneously for age, calendar year, and race; (2) fitting the same Poisson regression model, we employed the Firth [7] correction and the data augmentation method described by Greenland [8,9] to remove sparse-data bias; and (3) we ran simulations to investigate the extent of bias stemming from sparse data.

The Poisson regression model that we fit included terms for age, the square and cube of age, calendar year, the square and cube of calendar year, and race. Age and calendar year were centered at approximate median values (60 for age and 1975 for year) and were scaled in units of decades to avoid small coefficients, and then grouped into half-decade categories, each assigned a value that corresponds with the category midpoint. The RR results from this model are shown in the third column of Table 3. The estimate of exposure effect is smaller in this analysis than in any of the stratified analyses, with an estimate for the RR of 1.94 for the highest level of exposure relative to the lowest. Adding product terms between age and calendar year did not change the results substantially. Adding the Firth correction for sparse-data bias shifted the estimate of effect slightly more toward the null, to 1.91. Using data augmentation with one case per prior record induced greater shrinkage, giving a RR of 1.77 for the highest level of exposure relative to the lowest.

Our last set of analyses was intended to assess the amount of bias from sparse data by using simulations to separate the effect of sparse data from any effect of exposure. We randomly generated cases without reference to exposure level, using population mortality rates. We then analyzed the results of each simulation using the same stratification as in the NIOSH paper. Apart from the effect of biases introduced in the analysis, we would expect to find null results. The simulation results are shown in Table 4, which compares the results from the analysis of Schubauer-Berigan et al. for four selected exposure metrics with the respective medians of 1000 simulated analyses in which exposure was unrelated to lung cancer death. The simulations gave essentially null results for three of the four exposure metrics. For time since first employment, however, the simulations resulted in strong associations, with the median result comparing the highest level of exposure with the lowest being an SRR of 2.58. The departure of this result from 1.0 is a measure of the bias stemming from the stratified analysis. When the simulated data for time since first employment were reanalyzed

Table 1

Crude lung cancer counts, person-time rates and rate ratios by time since first employment

	Time since first employment (yrs)				Total
	0–15	15–25	25–35	≥ 35	
Cases	27	71	123	324	545
Persons	9199	8378	7449	6183	31,209
Person-years	131,388	79,481	68,459	73,190	352,518
Rate ($\times 10^5$ person-years)	21	89	180	443	155
Crude rate ratio	1.00	4.35	8.74	21.5	
SRR (NIOSH)	1.00	2.24	2.83	3.68	

NIOSH = National Institute of Occupational Safety and Health; SRR = standardized rate ratio.

Data from Schubauer-Berigan et al. [5].

Table 2
Lung cancer and person-time stratified by age and time since first employment

Age (yrs)		Time since first employment (yrs)				Total
		0–15	15–25	25–35	≥35	
10–24	Cases	0	0	0	0	0
	Person-years	14,538	0	0	0	14,538
25–34	Cases	0	0	0	0	0
	Person-years	48,823	2137	0	0	50,960
35–44	Cases	2	3	0	0	5
	Person-years	36,685	31,615	2069	0	70,369
45–54	Cases	10	12	22	3	47
	Person-years	19,375	26,505	30,243	1915	78,038
55–64	Cases	9	30	42	74	155
	Person-years	8862	12,977	23,684	26,047	71,570
65–74	Cases	5	23	50	143	221
	Person-years	2750	5068	9679	29,520	47,017
≥75	Cases	1	3	9	104	117
	Person-years	355	1179	2784	15,708	20,026
Total	Cases	27	71	123	324	545
	Person-years	131,388	79,481	68,459	73,190	352,518
Age standardized rate ($\times 10^5$ person-years)		73	134	139	195	
	SRR	1.00	1.82	1.90	2.65	
	RR _{MH}	1.00	1.72	1.83	2.73	

RR_{MH} = Mantel-Haenszel pooled estimator for rate ratio; SRR = standardized rate ratio.
Data from Schubauer-Berigan et al. [5].

using Poisson regression instead of stratification, the median estimate of the rate ratio over the 1000 simulated datasets was 1.00.

It is interesting that none of the other exposure metrics showed any indication of sparse-data bias. Consider employment duration. Schubauer-Berigan et al. reported SRRs for employment duration with a 10-year lag of 0.91, 0.92 and 1.03 for exposure categories 2 through 4, relative to the lowest exposure category. The simulations found null estimates for all levels of this metric. Unlike time since first employment, however, for which, comparing highest with lowest level of exposure, only 16% of person-time fell into strata with some person-time for both of the extreme exposure categories, for employment duration with a 10-year lag, the proportion of person-time falling into such strata was 85.9%. Similarly, for time since first employment, there were zero strata out of 448 that had nonzero case counts for both of the extreme exposure categories, whereas for employment duration with the 10-year lag, 34 strata had nonzero rates in both of the extreme exposure categories. These situations are substantially different. For cumulative exposure with a 10-year lag in subjects with over 1-year employment duration, the SRR for highest versus lowest exposure was 1.97, whereas the simulations found only null effects, revealing no sparse-data bias. For this exposure metric, 91% of the person-time fell into strata with some person-time for both exposure

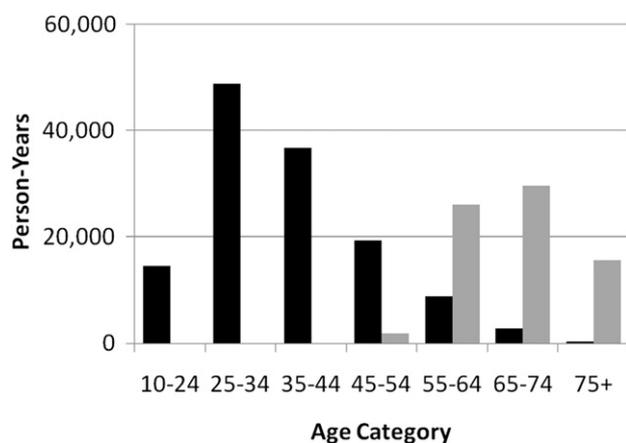


Fig. 1. Person-time distributions for lowest (black bars) and highest (gray bars) categories of time since first employment, by age, using seven age categories.

categories, and there were 23 strata with nonzero rates for both of the extreme exposure categories. This distribution of information was presumably sufficient to avoid sparse-data bias. The situation was similar for the other simulated analyses. Thus, the only exposure metric that showed the sparse-data bias was time since first employment. That was also the exposure metric that showed by far the largest SRR for fatal lung cancer in the NIOSH study.

Discussion

Greenland [6] suggested that “The simplest diagnostic for small-sample or sparse-data problems is close tabular examination of basic data.” Inspection of strata in a stratified analysis should be a routine preliminary step in data analysis [15]. For the data of Schubauer-Berigan et al., inspection of the data analyzing the effect of time since first employment reveals that workers with high levels of exposure and those with low levels of exposure had strikingly different distributions over the stratification variables, with overlap so slight that no stratum comparing the highest level of exposure with the lowest had a nonzero count of lung cancer cases for both of these exposure categories. When stratified data are that sparse, estimation of ratio measures is biased upward.

A comparison of the crude data and the result after stratified analysis indicated that there was strong age confounding in the

Table 3
Poisson regression model for examining time since first employment, with sparse-data corrections

Term	Coefficient	Standard error	RR	RR with Firth correction	RR with data augmentation
Intercept	−15.1885	0.3339			
TSFE 15–24 yrs	0.1181	0.2513	1.12	1.11	1.05
TSFE 25–34 yrs	0.2189	0.2679	1.24	1.23	1.14
TSFE ≥35	0.6610	0.2837	1.94	1.91	1.77
Age	1.0167	0.0849	2.76	2.75	2.76
Age ²	−0.3167	0.0574	0.73	0.74	0.73
Age ³	0.0163	0.0253	1.02	1.01	1.02
Year	−0.1145	0.0893	0.89	0.89	0.90
Year ²	−0.1707	0.0429	0.84	0.85	0.84
Year ³	0.0438	0.0159	1.04	1.04	1.05
Race (white/other)	−0.0518	0.2286	0.95	0.93	0.90

RR = rate ratio; TSFE = time since first employment.
Data from Schubauer-Berigan et al [5].

Table 4

Median SRR for 1000 simulations of lung cancer deaths in the NIOSH cohort based on no exposure effect, and SRR estimates from the NIOSH analysis, for four beryllium exposure metrics

Exposure metric	Exposure category			
	Lowest	Level 2	Level 3	Highest
Time since first employment lag 0				
NIOSH-reported results	1.00	2.24	2.83	3.68
Simulations	1.00	2.00	2.81	2.58
Duration employed lag, 10 yrs				
NIOSH-reported results	1.00	0.92	0.91	1.03
Simulations	1.00	1.00	1.01	0.99
Cumulative exposure lag, 10 yrs				
NIOSH-reported results	1.00	1.56	1.66	1.97
Simulations	1.00	0.98	1.00	1.01
Maximum exposure lag 0				
NIOSH-reported results	1.00	1.83	1.89	1.50
Simulations	1.00	1.01	1.00	0.98

NIOSH = National Institute of Occupational Safety and Health; SRR = standardized rate ratio.

NIOSH data. That is not surprising, inasmuch as age is a strong risk factor for lung cancer and is strongly correlated with time since first employment. The crude effect estimate of time since first employment, comparing highest with lowest categories, was 21.5. Simply stratifying into seven age categories gives a much different result, a standardized RR of 2.65, revealing a confounding effect of one order of magnitude, a level of confounding that is seldom seen in real data. The other stratification variables in the NIOSH analysis—calendar time and race—were both confounding in the same direction based on univariate analyses, although not as strongly as age. One would expect that when control of the additional confounders was implemented, the effect estimate would get even closer to the null than the value of 2.65 obtained from the control of age alone. In fact, however, the estimate increased to 3.68 with control of the additional two variables and the finer age stratification.

This shift toward a higher rather than a lower value can be explained by sparse-data bias stemming from the small numbers and lack of overlap between compared exposure categories after the very fine stratification called for when all three confounding factors were controlled simultaneously. As our simulations demonstrated, the sparse-data bias in these data can be removed by using a Poisson regression model to analyze the data; for our simulated data that had a null relation between exposure and lung cancer, a biased estimate of 2.58 was the median result of the stratified analysis for the highest level of exposure, whereas the corresponding median result from Poisson regression analysis was 1.00, indicating no bias. When we used Poisson regression to analyze the NIOSH data for this same contrast, we obtained a value of 1.94 from Poisson regression, and slightly lower using the Firth correction or the data augmentation method of Greenland. These are more plausible estimates than the originally reported 3.68 for control of all three confounders, given that all three variables confound in the same direction and that control of age alone led to a SRR of 2.65 (see Table 2). If the difference between 1.77 from the data augmentation method and the value of 3.68 reported in the NIOSH paper were taken as an indicator of the amount of sparse-data bias present, it implies that $(1.77-1)/(3.68-1) \approx 29\%$ of the effect estimate originally reported represents the actual association with exposure, and 71% of the effect estimate stems from sparse-data bias.

We note that a full re-analysis of these data could take advantage of the quantitative information in the exposure metrics by avoiding categorization of the exposure. Modeling the effect of a continuous exposure metric could be coupled with spline

regression to fit a smooth curve describing the relation between exposure and outcome that is not hampered by strong modeling assumptions.

The data under discussion are the focal data in the literature used to quantify the relation between beryllium and lung cancer. Changes in work exposures make it unlikely that these data will be supplanted by future studies. Therefore, we believe that detailed re-analyses such as we present are critical to help decipher the actual magnitude and nature of the relation between beryllium and fatal lung cancer. This re-examination also serves as a useful reminder about potential pitfalls in data analysis that need to be kept in mind by researchers in the course of conducting everyday research.

Acknowledgments

This study was funded by a contract between Materion Corporation and RTI Health Solutions. RTI Health Solutions, a business unit of the Research Triangle Institute, is an independent, nonprofit, research organization that does work for government agencies and private companies. RTI had full control over the content of this work. The authors are grateful to NIOSH for providing us a copy of the data used by Schubauer-Berigan et al. in their 2011 publication. We are also grateful to Sander Greenland for his review and advice.

References

- [1] Rothman KJ, Mosquin PL. Confounding after risk-set sampling in the beryllium study of Sanderson et al. *Ann Epidemiol* 2011;21:773–9.
- [2] Sanderson WT, Ward EM, Steenland K, Petersen MR. Lung cancer case-control study of beryllium workers. *Am J Ind Med* 2001;39:133–44.
- [3] Deubner DC, Lockey JL, Kotin P, et al. Re: Lung cancer case-control study of beryllium workers. Sanderson WT, Ward EM, Steenland K, Petersen MR. *Am J Ind Med* 2001;39:133–144. *Am J Ind Med* 2001;40:284–5.
- [4] Schubauer-Berigan MK, Deddens JA, Steenland K, et al. Adjustment for temporal confounders in a reanalysis of a case-control study of beryllium and lung cancer. *Occup Environ Med* 2008;65:379–83.
- [5] Schubauer-Berigan MK, Couch JR, Petersen MR, et al. Cohort mortality study of workers at seven beryllium processing plants: update and association with cumulative and maximum exposure. *Occup Environ Med* 2011;68:345–53.
- [6] Greenland S. Small-sample bias and corrections for conditional maximum-likelihood odds-ratio estimators. *Biostatistics* 2000;1:113–22.
- [7] Firth D. Bias reduction of maximum likelihood estimates. *Biometrika* 1993;80:27–38.
- [8] Greenland S. Bayesian perspectives for epidemiological research. II. Regression analysis. *Int J Epidemiol* 2007;36:195–202.
- [9] Greenland S. Variable selection versus shrinkage in the control of multiple confounders. *Am J Epidemiol* 2008;167:523–9.
- [10] Steenland K, Spaeth S, Cassinelli R 2nd, et al. NIOSH life table program for personal computers. *Am J Ind Med* 1998;34:517–8.
- [11] Schubauer-Berigan MK, Raudabaugh WR, Ruder AM, et al. LTAS.NET: a life table analysis system for the windows environment [abstract]. *Ann Epidemiol* 2005;15:656.
- [12] Schubauer-Berigan MK, Hein MJ, Raudabaugh WM, et al. Update of the NIOSH Life Table Analysis System: a person-years analysis program for the Windows computing environment. *Am J Ind Med* 2011;54:915–24.
- [13] Rothman KJ. *Epidemiology, an introduction*. 2nd ed. New York: Oxford University Press; 2012.
- [14] Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* 1968;24:295–313.
- [15] Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott, Williams & Wilkins; 2008.

Appendix

Simulation of Cohorts

Historical datasets for the NIOSH study [5] were obtained from NIOSH. Following NIOSH, the only female and 11 subjects who lacked dates of birth were removed from the analysis, leaving 9199 subjects. Our data differed from the NIOSH data in being

de-identified for date of birth and date last observed, with the day for each date being suppressed. Values for these suppressed days were imputed as the 15th of the month.

We generated 1000 simulated cohorts. Within each, individual dates and causes of death were generated using rates from *LTAS.NET* [11]. Starting with their first year of employment and continuing for each year thereafter up to 2005, we simulated whether death occurred by lung cancer or other causes, where the probability and cause of death depended only on the age-specific *LTAS* rate and time at risk. In the simulation, mortality does not depend on beryllium exposure.

Overall mortality rates for White males for lung cancer and all other causes were obtained from NIOSH *LTAS.NET* with the rates for other causes set equal to the sum of all rates for non-lung cancer causes. The NIOSH rates apply to 5-year time periods and 5-year age ranges (e.g., ages 15–19, years 1960–1964). We assigned these rates to each of the 25 single age–year combinations represented by a single NIOSH rate for grouped age and year data, here represented as r_{jk}^L for lung cancer and r_{jk}^O for other causes, for individual year

$j = 1940, \dots, 2005$ and age $k = 14, \dots, 120$. Subjects hired at age 14 were assigned the age 15 rates and subjects age 85 and older the age 85+ rates.

Mortality was simulated within discrete time intervals starting with the first day of employment. For a person alive at the start of an interval, the interval extended either to their next birthday or the end of the calendar year, whichever occurred earlier. For each interval, two independent random uniforms $0 < u_L < 1$, $0 < u_O < 1$ were generated, with death occurring by lung cancer if $u_L < p_L$ and $\frac{u_L}{p_L} \leq \frac{u_O}{p_O}$ and by other causes if $u_O < p_O$ and $\frac{u_L}{p_L} > \frac{u_O}{p_O}$. Quantities p_L, p_O are functions of the length of the interval T and the specific year j and age k associated with the interval. For example, for lung cancer $p_L = 1 - \exp(-(\log(1 - r_{jk}^L)T))$ and p_O is found similarly. If the cause of death is lung cancer, age at death is set equal to age at the start of the interval plus $\frac{Tu_L}{p_L}$ and $\frac{Tu_O}{p_O}$ if death is from other causes. For each person, simulation continued to mortality or December 31, 2005, whichever came first.