ADJUSTMENT FOR SUSPECTED MISCLASSIFIED SMOKING DATA IN AN HISTORICAL COHORT STUDY OF WORKERS EXPOSED TO ACRYLONITRILE

by

Sarah Downing Zimmerman

B.S., Mathematics, John Carroll University, 2010

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Sarah Downing Zimmerman

It was defended on

November 13, 2013

and approved by

Ada Youk, PhD, Assistant Professor, Department of Biostatistics,

Graduate School of Public Health, University of Pittsburgh

Evelyn Talbott, PhD, Professor, Department of Epidemiology,

Graduate School of Public Health, University of Pittsburgh

Jim Collins, PhD, Adjunct Professor, Department of Biostatistics,

Graduate School of Public Health, University of Pittsburgh

Thesis Director: Gary Marsh, PhD, Professor, Department of Biostatistics,

Graduate School of Public Health, University of Pittsburgh

ADJUSTMENT FOR SUSPECTED MISCLASSIFIED SMOKING DATA IN AN HISTORICAL COHORT STUDY OF WORKERS EXPOSED TO ACRYLONITRILE

Sarah Downing Zimmerman, MS

University of Pittsburgh, 2013

ABSTRACT

Objectives: To examine the association between exposure to acrylonitrile (AN) and lung cancer mortality after properly addressing misclassification and possible positive confounding of smoking history.

Methods: Subjects were 992 white males who were employed for three or more months between 1960 and 1996 at an AN chemical plant in Lima, OH. There were 15 identified cases of lung cancer deaths. Smoking histories were obtained for 90.3% of the cohort and 54.2% of the cohort were identified as having "ever smoked". Though there were few "unknown" smoking histories, the smoking variable was determined to be misclassified as the RR for having ever smoked related to lung cancer was only 1.08 (95% CI=0.26, 6.18). We addressed potential confounding by smoking in the presence of suspected misclassified smoking data by determining if a reasonable adjustment of the available smoking data would change the risk levels of lung cancer in the original Lima cohort and the relationship between AN exposure and lung cancer using Monte Carlo simulation and bias adjustment.

Conclusions: After running Monte Carlo simulation, we found that the mean RR of lung cancer mortality given differing levels of AN exposure decreased after adjusting for the simulated smoking data. However, the results from the bias adjustment must be interpreted with caution as the analysis

was limited by the number of lung cancer cases. In this cohort, we concluded that smoking positively confounded the relationship between AN exposure and lung cancer mortality.

Public Health Relevance: Properly adjusting for smoking history in studies of lung cancer is critical of the validity of the study results. As seen in this study, smoking habits impact the risk of certain health outcomes. Researchers must attempt to address the potential confounding by smoking whenever possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to specially thank my thesis committee for your patience and guidance. This thesis could not have been written without the help of Dr. Gary Marsh, who challenged and encouraged me throughout my academic program. I could not have completed my statistical analysis without Dr. Ada Youk, who patiently taught me how to successfully run OCMAP. Thank you to Dr. Evelyn Talbott and Dr. James Collins who presented insightful critiques of my thesis. None of you accepted less than my best efforts. Thank you very much for helping me in furthering my education.

I would also like to thank my family who supported my education, my choice to study in biostatistics, and gave thousands of words of encouragement. Lastly, thank you to my husband, Scott Zimmerman, who has never stopped supporting or encouraging me in my studies. You have helped me in uncountably many ways, and I am so very thankful to you.

1.0 INTRODUCTION

1.1 BACKGROUND AND RATIONALE

Acrylonitrile (AN) is a chemical used in the production of many plastics, synthetic fibers and rubbers, and previous experimental studies have shown AN to be carcinogenic in animals. Excess exposure to AN in rats has resulted in certain types of cancer (astrocytomas), occurring in the brain and spinal cord, and tumors of the Zymbal gland, forestomach, stomach and mammary gland (Strother et al., 1998). Due to the prevalence of AN in manufacturing, many studies focus on human exposure to the chemical in occupational settings, and researchers are interested in determining if AN is potentially carcinogenic to humans. To date, epidemiologic studies have provided inconclusive evidence to support a claim that AN is carcinogenic to humans. AN is currently labeled as a "possible" (Group 2B) carcinogen by the International Agency for Research on Cancer (IARC, 1999).

There are four main occupational cohort mortality studies that focus on the health effects of AN. The National Cancer Institute (NCI) performed the largest study to date which analyzed the relationship between the manufacture and use of AN and mortality rates among employees in eight facilities (Blair et al., 1998). In this study, researchers attempted to find a relationship between

varying levels of exposure to the chemical and changes in mortality rate of the workers. They concluded that the exposure levels in the facilities did not create a significant increase in relative risk of cancer deaths, but they did notice an "excess of lung cancer in the highest quintile of cumulative exposure." Similar analyses were performed in other cohort studies, for example in South Carolina (DuPont cohort- Symons et al., 2008), the UK (Benn et al., 1998), and the Netherlands (Swaen et al., 1998). These studies reported slight elevations in risk for several types of cancer mortality, including lung cancer, but found no statistically significant relationship between exposure to AN and the risk of lung cancer mortality.

Researchers would like to conduct a full evaluation of this increased risk for lung cancer, however they are limited by the absence or misclassification of smoking data. In all studies of lung cancer, one must take into account the effects of smoking on the occurrence of lung cancer as there is a well-known relationship between these factors. However, in the UK, Dutch, and DuPont analyses mentioned above, smoking data are missing, and it is possibly misclassified in the NCI study. As such, one cannot draw full conclusions about the association between AN exposure and lung cancer in these studies.

In 1995, the University of Pittsburgh's Center for Occupational Biostatistics and Epidemiology (COBE) was commissioned by BP Chemicals, Inc. (BPC) to perform an updated and extended investigation of mortality patterns with an emphasis on cancer mortality in relation to AN exposure within their chemical plant located in Lima, OH (Marsh et al. 1999), which was one of the eight plants studied in the NCI cohort study. At BPC's request, COBE updated the mortality and work histories of employees from this plant to include data on those employees who worked at the plant

for at least three months between 1960 and 1996. The study gathered data on 992 white males who worked in the plant during this time period. Of these 992 workers, 15 of them died from lung cancer. In this update, smoking histories were also provided to enable an evaluation of the relationships among AN exposure, lung cancer mortality, and smoking history.

In the Marsh et al. publication (1999), there was limited evidence of an exposure-response trend between AN exposure and lung cancer after adjusting for time since first employment. In the lowest (>0 - 4.9 ppm), middle (5.0 - 11.9 ppm) and highest (12.0+ ppm) categories of average intensity of AN exposure, the relative risk (RRs) and associated confidence intervals were 1.18 (95% CI=0.16, 6.84), 1.46 (95% CI=0.22, 7.29) and 2.91 (95% CI=0.46, 14.13) respectively. They concluded that, although the exposure-response analysis showed monotonically increasing RRs for lung cancer given AN exposure, this trend was not statistically significant. Accurately accounting for smoking history in the evaluation of the relationship between AN exposure and lung cancer is a necessary step in evaluating an exposure-response trend.

In the Marsh et al. study (1999), however, the final exposure-response results were not adjusted for smoking histories of the workers because the RR for lung cancer mortality related to smoking history was inordinately low. This RR was 1.08 with 95% confidence interval (0.28, 6.18). It is well known (as discussed later) that smoking is a major risk factor of lung cancer, and this relationship was not evident in the smoking data from the 1999 study. As the smoking variable was not a statistically significant predictor of lung cancer mortality, the data was considered to be highly misclassified, and thus smoking information was not included in the analysis. This thesis will address the issue of possible smoking misclassification in the 1999 study.

1.2 STATEMENT OF THE PROBLEM

In this thesis, using Monte Carlo simulation and bias adjustment, I will address potential confounding by smoking in the presence of suspected misclassified smoking data by determining if a reasonable adjustment of the available smoking data changes the risk levels of lung cancer in the original Lima cohort and the relationship between AN exposure and lung cancer. This will provide a more valid assessment of the association between AN exposure, lung cancer mortality, and smoking history. The focus of this thesis will use the Lima, OH cohort updated by COBE (Marsh et al., 1999) with the study period January 1, 1960 to December 31, 1996.

1.3 RELATIONSHIP BETWEEN SMOKING AND LUNG CANCER

Smoking history for the workers in the original Lima cohort was determined by a voluntary mail-in survey issued by BP. Also, medical records were examined by the original researchers to obtain more information on lifetime smoking history. The relationship between smoking history and AN exposure is displayed in Table 1. The first column groups workers based on their cumulative exposure to acrylonitrile over the course of their entire employment at the plant. For example, among of all workers with a cumulative exposure over 110 parts per million per year (ppm-years), three workers died of lung cancer, and 61 of the workers in this category admitted to having ever smoked (which accounts for 80.3% of all people in this category). Also, 13 workers with exposure rates above 110 ppm-years claimed to have never smoked (17.1% of the category). Cumulative exposure is displayed in this table (rather than duration of exposure or average exposure) to enable the comparison of smoking histories of those in the highest exposure levels to those unexposed to AN. Within the total cohort, smoking data were gathered for 90.3% of workers and 93.0% of AN

exposed workers. Further, 54.2% of the workers were identified as having "ever smoked". Those without reported smoking histories were classified as unknown.

Table 2 displays results from Marsh et al's (1999) univariate risk estimates for smoking history, AN duration (Dur) of exposure, AN cumulative (Cum) exposure, and AN average intensity of exposure (AIE). In the row corresponding to "Smoking History", we note that, of the 15 reported deaths due to lung cancer, three people claimed to have never smoked, 10 claimed to have smoked and two had unclassified smoking data. Here, it is shown that the univariate RRs of lung cancer mortality due to smoking is 1.08 with associated p-value=0.999, while the RR of those with unknown smoking data is 1.18. Because the RR of lung cancer mortality due to smoking history was close to 1.00 for both ever and unknown smokers in the original study, smoking history was not a significant main effect of lung cancer mortality, and was thus not included in the final statistical model developed by Marsh et al. (1999). Further explanation of relative risk regression is in Section 2.1.

Table 3 shows a summary of the relative risk regression analysis for lung cancer mortality adjusted for time since first employment as reported by Marsh et al. in 1999. This thesis will investigate whether reassigning smoking history data through simulation will change the exposure-response trend reported by Marsh et al. by comparing the adjusted data to the data displayed in Table 3.

Earlier studies have shown that the risk of lung cancer increases 11-fold (Higgins and Wynder, 1988) among those who have ever smoked and reported the odds ratio to be 35.5 (Samet, 1993) among heavy smokers compared to non-smokers. The US Surgeon General's report (2006) shows a comparison of lung cancer standardized mortality ratios (SMR) between smokers and

nonsmokers. The reported SMRs for smokers ranged from 2.03 to 14.20 in a Japanese study (Loeb et al., 1984) and a Canadian Veterans study (Loeb et al., 1984), respectively. The American Cancer Society (Loeb et al., 1984) reported the lung cancer SMR to be as high as 10.73 in smokers compared to nonsmokers.

Given the RRs for lung cancer mortality due to smoking observed in the literature, the "inordinately low" RR mentioned above of 1.08 reported by Marsh et al. (1999) strongly suggests smoking was misclassified among members of this cohort. In the first part of this thesis, we attempted to determine if adjusting the data for smoking and then including it in the final analysis would change the results (thereby showing that misclassification of smoking data has a strong impact on the outcome of the study). This was done by randomly reassigning a percentage of workers labeled as "non-smokers" to "smokers" repeatedly via a Monte Carlo process, and repeatedly analyzing the results. This process generated a more realistic RR between the smoking data and the lung cancer mortality cases. In the second part of this thesis, we attempted to remove any bias related to smoking history from the final RR of lung cancer mortality related to AN exposure using Richardson's method (2010) described in more detail later.

1.4 CONFOUNDING BY SMOKING

Confounding can occur when a third variable (here, smoking history), which is not in the causal pathway between the exposure and health outcome (here, AN exposure and lung cancer), is related to both factors. The first step is to determine if the potential confounder is a statistically significant risk factor for the health outcome. If it is a risk factor, then the variable (smoking history) has the

potential to confound the exposure-response relationship if it is also related to the exposure variable (AN exposure). It is possible that the limited evidence of a relationship between AN exposure and lung cancer in the NCI study (Blair et al., 1998) and the Lima cohort study (Marsh et al., 1999) was due, at least in part, to a misclassification of smoking history among the workers, and a more accurate assessment of the association (or lack thereof) will result after accounting for the potential confounding by smoking history.

A well-known example of confounding occurred in a classic study in which researchers were attempting to find causes of lung cancer. In this study (Marsh et al., 1988), researchers noticed that many of the participants who had lung cancer also frequented bars. From this information, they drew the conclusion that alcohol increased the risk for lung cancer. However, the researchers failed to take into account that these participants were inhaling a large amount of cigarette smoke while at the bar. This is an example where the confounding variable (smoking) affected the conclusion relating the risk factor (alcohol) and health outcome (lung cancer). Specifically, this is an example of positive confounding as the relationship between alcohol consumption and lung cancer was artificially increased by the confounding effects of smoking. Negative confounding occurs when the confounding variable masks the association between the risk factor and health outcome.

If smoking is more prevalent among workers with higher exposure rates and less prevalent among those with lower rates, positive confounding by smoking is likely to have occurred between these variables. In the Lima cohort study (Marsh et al. 1999), those workers who experienced a higher cumulative exposure to AN were more likely to report having ever smoked. This relationship is evident in Table 1; as AN exposure increases, the number of those identified as “ever smoked” is

larger than “never smoked”. Thus, positive confounding by smoking is a possibility in this study. In this thesis, we attempted to identify and account for the extent to which smoking was positively confounding the relationship between AN exposure and lung cancer mortality.

As shown in Table 2, there was a monotonically increasing exposure-response trend for AN exposure and lung cancer mortality, and a sufficient amount of smoking data is present to analyze the effect of confounding on the results. We were able to adjust the smoking data through simulation and investigate the resulting effect on the exposure-response relationship.

2.0 PROPOSED METHODS FOR THESIS

2.1 OVERVIEW OF METHODS

Several approaches are available for adjusting an exposure-response analysis for a potential confounding variable. This thesis will implement and compare two techniques to account for the potential confounding effect of smoking on the relationship between AN exposure and lung cancer mortality.

The first technique used to adjust for confounding is a sensitivity analysis using the Monte Carlo (MC) method. The MC method is a general procedure that is used to run simulations many times. In this thesis, we will simulate the problematic variable (i.e. smoking history) by random reassignment of the “never smoked” label to “ever smoked”. We will then perform regression analysis using the

new variable and repeat this combination using the simulated variable many times. Using the MC method, we can analyze the sensitivity of the RRs and thereby answering the question "how much did these RRs change as a proportion of the smoking data was changed from never smoked to ever smoked"?

Steenland and Greenland used a very similar method in their publication, "Sensitivity Analysis of an Unmeasured Confounder" (2004), where the authors used and compared two different methods to adjust for confounding by smoking on lung cancer risk: MC sensitivity analysis and Bayesian bias analysis. Steenland and Greenland concluded that the two methods yielded similar results and that these types of analyses "should be more widely adopted by epidemiologists" (2004). Also, a similar approach was used by Cunningham (2005) in his unpublished thesis to adjust for missing smoking data in the NCI case-cohort study. Though similar in methodology, this thesis will address misclassification rather than missing data in smoking history using the MC approach.

The second approach, developed by Richardson (2010), is used if smoking history is too difficult to obtain or if there exists a large proportion of missing data. This is a common issue, as it is rare to accumulate accurate smoking data in a cohort study. Richardson proposed a method in which one can estimate and remove the bias of smoking from the RR of AN exposure related to lung cancer mortality. Richardson suggests using this method as this approach does not require smoking data or assumptions of the smoking variable distribution.

2.2 FIRST METHOD: MONTE CARLO (MC) SIMULATION

The following general steps are required to incorporate the Monte Carlo (MC) approach:

Step 1: Adjust data to achieve a realistic odds ratio (OR) for smoking and lung cancer similar to those shown in the Surgeon General's Report (2010),

Step 2: Generate risk sets using OCMAP (Marsh et al 1998), a statistical software, to help adjust for the effect of potential confounding due to age and time period on health outcome,

Step 3: Simulate a new smoking history variable using the MC approach,

Step 4: Run conditional logistic regression models to estimate the relative risk (RR) of smoking using the simulated smoking variable,

Step 5: Compare the new RRs to the original RRs to determine if adjusting for the new smoking variable affects the risk of lung cancer given exposure to AN.

Step 1: Adjust Data to Achieve a More Realistic Odds Ratio

According to the Surgeon General's Report (1986, 2010), about 90% of lung cancer cases can be attributed to smoking. In the original Lima cohort, only around 67% of those people who died of lung cancer were identified as smokers according to the literature. This, along with the inordinately low RR for smoking and lung cancer mortality observed in the original cohort by Marsh et al. (1999), provides evidence that the smoking data are possibly misclassified, as a higher percentage of those who died of lung cancer should have also been smokers. In order to achieve a more

realistic percentage, the smoking data must be reclassified through simulation to reflect a higher smoking prevalence within lung cancer deaths.

Table 4 illustrates how the odds ratio (OR) of lung cancer mortality related to smoking increases as the prevalence of smoking is increased. The ORs were calculated using 2x2 tables. The first row in the table displays the information about the original data set. In this set, 10 workers who died of lung cancer (cases) were identified as having ever smoked, and five of the cases were identified as never having smoked or had missing smoking data. Thus, about 66.7% of all cases had identified themselves as smokers. The OR for this scenario is presented in the fifth column, and the last column displays the p-value from a Fisher's Exact Test on the null hypothesis that the OR equals 1.00. The following rows display scenarios in which the prevalence of smoking was increased by a value of one each time. In other words, one case labeled as "never smoked" or "unknown" was reassigned to "ever smoked." In each of these new scenarios, the smoking prevalence, OR, and p-value were calculated. These changes were made with the purpose of adjusting the smoking prevalence among cases and increasing it to a more realistic value similar the one reported by the Surgeon General (1986, 2010).

Table 4 shows that as the smoking prevalence increased within cases, the OR increased as well. As the prevalence approached 90.0%, the OR approached 10, which is expected as the Surgeon General's report (1986, 2010) concluded that the SMR for lung cancer given smoking is around 10. In the final and most extreme scenario, the p-value is 0.028 indicating a statistically significant relationship between smoking and lung cancer mortality exists.

This thesis will focus on the results shown in Table 4 Scenario 4. Recall that, in order for a variable to cause confounding in a study, a relationship between the confounding variable and the main risk factor is necessary. This was shown to be the case between AN exposure and smoking in this data set in Table 1. Additionally, a statistically significant relationship between the confounding variable and the health outcome must be present. As such, we will be able to determine if smoking confounds the relationship between AN exposure and lung cancer death.

In this analysis, we assumed that the misclassification is biased downward away from "ever smoking" rather than biased toward "never smoking". One possible explanation for this misclassification is the self-conscious response by workers being interviewed for their medical records. Considering that there is a social stigma associated with being identified as a smoker, we will assume that many individuals would misrepresent themselves as a nonsmoker when, in reality, they have smoked. Workers may be unlikely to admit that they smoked when being checked by a health professional during regular checkups, and they may also be unwilling to admit to the company that they smoke due to negative responses from health insurance providers. For these reasons, we chose to focus on an increased prevalence of smoking within the simulation to account for the supposedly low amount of smoking data.

Step 2: Generate Risk Sets

The technique of creating a risk set is a method which matches employees who died of lung cancer (cases) with those who did not (controls) based on some specific criterion. A risk set was created for each case by matching the controls who were alive at the exact age at which the case died. That

is, if a case died at the age of *x*, then all controls who were at risk and lived to be at least *x* years old were grouped with this case. The risk set was then further matched on year of birth using caliper matching by only considering those controls whose birth date was within some determined range of the case. Thus, the case that died at *x* years old was matched with all controls that lived to be at least *x,* were at risk and were born within five years before or after the case's birthday. The range of birthdates around each case was originally one year, but this created risk sets that were too small. As such, the range was increased to five years for everyone, and ten years for two cases (in order to create large enough risk sets). To properly analyze the data, the statistical software, OCMAP (Marsh et al. 1998), was used to create these risk sets for each case. By the method with which these risk sets were matched, it is possible that some controls appear in multiple risk sets, and some do not appear in any of them.

Creating risk sets helps eliminate the confounding effects caused by the matched variables which, in this instance, are exact age and time. Age is a major factor in the onset of cancer, therefore, if controls and cases are matched based on exact age at death, confounding due to age should be reduced during the analysis within risk sets. Additionally, matching on year of birth helps eliminate any birth cohort effects.

Table 5 displays the counts for each risk set partitioned by smoking history information. Each row represents a different risk set for each case as described above. There are 15 risk sets as there were 15 cases of lung cancer in the data set. The second column displays the number of controls matched with each case. The following three columns each contain a pair of numbers indicating the number of cases and controls in that risk set labeled at the listed level of smoking. For example, in risk set

one (which contains case one), 22 controls were matched, and column three shows that the case was identified as having ever smoked while three controls were labeled in this way. Column 4 shows that 16 controls in this risk set were labeled "never smoked", and the last column indicates that three controls had unknown smoking data.

Step 3: Simulate New Smoking Variable

As mentioned above, the remainder of the analysis and simulation was conducted using adjusted data from Scenario 4 in Table 4 (which contained the most realistic OR for lung cancer and smoking prevalence among cases). In an attempt to remove confounding from the likely misclassification of the smoking data, we simulated new smoking history data using the statistical software STATA (StataCorp, 2005). The STATA code is shown in Appendix B. In these simulations, we reassigned some of the nonsmokers and all of those with missing smoking data to "ever smoked", and then statistical analyses were performed with this new smoking information. The purpose of this simulation and reanalysis was to mimic the data adjustment from Table 4 on a large scale; we wanted to reassign the controls in a similar fashion to the cases. By comparing the results of the new analysis to the original results of the study, we would hopefully be able to determine if the smoking data confounded the association between AN exposure and lung cancer mortality.

The details of the simulation process are as follows. First, we reassigned all of the controls who had unknown smoking data as "never smoked." Then, every control labeled as "never smoked" had a 50% probability of being randomly reassigned as "ever smoked". Those controls that were not

chosen to be reassigned remained in the "never smoked" category. As for the cases, 10 of the total 15 were identified as "ever smoked" while five cases were labeled as either "never smoked" or "unknown." Similar to the controls, all cases listed as "unknown" were transformed to "never smoked." Next, a total of four of the five smokers were randomly chosen and reassigned as "ever smoked" so that the distribution of smokers from Scenario 4 in Table 4 was reproduced. Doing so resulted in 14 cases labeled as "ever smoked" while one case remained as "never smoked". This process of reassigning smoking histories was simulated 500 times and is referred to as "Scenario 4A". We chose to run this simulation 500 times as this provided enough estimates to create a stable mean of the RRs. Increasing the number of simulations did not change the mean RR substantially. This repeated simulation of data is known as a Monte Carlo simulation.

To create more extreme scenarios, we reassigned the smoking data based on the AN exposure category. In scenario 4A, we assumed no relationship between AN exposure and smoking as the reassigning of nonsmokers to smokers was done randomly and did not take AN exposure into account. However, as discussed earlier and seen in Table 1, employees with higher levels of AN exposure were more likely to be classified as smokers. Therefore, we simulated scenarios 4B and 4C to reflect this relationship. The rate at which controls were reassigned from "never smoked" to "ever smoked" was based on exposure category. In scenario 4B, "never smoked" controls with the highest average intensity exposure levels (higher than 12.0 parts per million per year) were reassigned as "ever smoked" 80% of the time while those with the lowest AIE levels (greater than 0 and less than 4.9 ppm-years) were reassigned only 30% of the time. In scenario 4C, these probabilities were 95% and 5% respectively. Average intensity exposure was chosen to represent

exposure levels rather than cumulative exposure because changes in AIE exhibited the strongest increasing trend in RRs for lung cancer (as seen in Table 3).

Step 4: Conditional Logistic Regression

Once we had the simulated smoking variables for scenarios 4A, 4B, and 4C, RRs were estimated by running conditional logistic regression models. As the outcome is binary (case or control) and the observations are matched within each risk set, modeling using conditional logistic regression is appropriate. This model is shown here:

$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = \widehat{\alpha_1} + \widehat{\alpha_2}x_2 + \cdots + \widehat{\alpha_{15}}x_{15} + \widehat{\beta_{16}}x_{16} + \widehat{\beta_{17}}x_{17} + \widehat{\beta_{18}}x_{18}.$$

In the Lima study, $\hat{p}$ was the probability that a worker died from lung cancer. In conditional logistic regression, the variables $x_2, x_3, \ldots, x_{15}$ represent which risk set a worker is assigned to through the matching process. Recall, all controls are matched to a case using the exact age at which the case died and further on the case's year of birth. AN exposure, time since first employment, and the simulated smoking history variables were the explanatory variables in the model. These are represented by x_{16}, x_{17}, and x_{18} respectively. The value of $\ln\left(\frac{\hat{p}}{1-\hat{p}}\right)$, or the logit, represents the natural log of the probability that an individual will be a case divided by the probability that they will be a control given specific variables. By exponentiating the outcome of a logistic regression, we obtain a value of the RR. Mathematically,

$$RR = exp\left(\ln\left(\frac{\hat{p}}{1-\hat{p}}\right)\right).$$

In this study, the RR may be interpreted as the probability that a workers died of lung cancer divided by the probability that one did not given the influence of certain variables. A higher RR

indicates that the given conditions are more likely to result in cancer death, while an RR close to 1.00 indicates that the risk of dying from lung cancer was not affected by variables in the model.

Step 5: Analysis of Relative Risk (RR)

The next step was to compare the simulated RRs to the original RRs from Marsh et al. (1999) (Table 3) to determine if adjusting for the new smoking variable affected the risk of lung cancer mortality related to AN exposure. These regression results are shown in Table 6. For comparison purposes, the model results in the top half of the table are taken from the original analysis by Marsh et al. (1999). The first section shows the original RRs for lung cancer given the two different options for smoking history, along with their confidence intervals; the next section displays the RRs of lung cancer given the four different levels of AN exposure adjusted for time since first employment (as seen in Table 3) and their associated confidence intervals.

The bottom half of Table 6 contains information regarding the results of the 500 simulations scenarios 4A, 4B, and 4C. The mean RR among all 500 simulations in scenario 4A (50% chance of "never smoked" controls reassigned as "ever smoked") for those who ever smoked was 1.48 with standard deviation 0.43, minimum 0.48 and maximum 2.96. Similarly the mean, standard deviations and extreme values of RRs for the four different levels of AN exposure can also be seen in this section of the table. This scenario produced mean RRs nearly identical to the original AN model from row two. Although the RR for smoker vs. nonsmoker increased from 1.08 to a mean of 1.48, the RRs for the exposure categories are very similar and the same increasing trend exists. This is

probably due to the fact that any prior relationship between AN exposure and smoking history was ignored during the simulation process in scenario 4A.

In scenario 4B (detailed results can be seen in following section), the mean RR for "ever smoked" increased to 2.48 which indicates a slightly more realistic relationship between smoking and lung cancer, and the mean RRs for each exposure category decreased marginally compared to the original RRs. Similarly, in scenario 4C, the mean RR for "ever smoked" increased to 2.92, and the mean RRs for the exposure categories decreased even more than in scenario 4B.

This relationship between mean RRs is shown more clearly in the box plot in Figure 1. There is a drop in the mean RR for exposure levels from the original model and Scenario 4A to Scenario 4C. Notice also that the increase in mean RR for the smoking data between the scenarios mirrors the drop in mean RR for the highest exposure level. It is possible that such a relationship between increased smoking RRs and decreased exposure RRs would be more evident with a more extreme adjustment of the data (to create smoking RRs similar to the well-known values), but such adjustment is not possible with such a small number of cases. This change in RRs is evidence for confounding by smoking: as the risk between smoking and lung cancer mortality increases, the risk between AN exposure and lung cancer mortality decreases.

2.3 SECOND METHOD: RICHARDSON'S METHOD

In this second method, we considered a health outcome that was highly correlated with the confounding variable (smoking) and not correlated with acrylonitrile exposure. Richardson

proposed to use the occurrence of chronic obstructive pulmonary disease mortality, or COPD, as the new variable. As there is no known relationship between COPD and AN exposure and a strong relationship between COPD and smoking, Richardson determined any observed relationship between COPD and exposure was solely due to a relationship between exposure and smoking. This relationship will be used to estimate a bias which will then be removed from the RR for lung cancer due to exposure and thereby eliminating any confounding due to smoking. The final result will hopefully show the true association between AN exposure and lung cancer.

The following are general steps required to incorporate Richardson's method:

Step 1: Using the original data, perform Richardson's method to estimate the bias due to confounding by smoking,

Step 2: Adjust the original RR for lung cancer due to AN exposure by the bias estimate,

Step 3: Analyze the results of Richardson's method.

Step 1: Estimate Bias due to Confounding by Smoking

Richardson developed a different method to estimate the bias due to confounding in his publication, "Occupational Exposures and Lung Cancer: Adjustment for Unmeasured Confounding by Smoking". In this paper, he outlined his method for adjusting RRs to account for unmeasured or unknown individual smoking history. As mentioned earlier, we will investigate the relationship between exposure and COPD as any association between exposure and COPD should reflect the relationship between exposure and smoking. If we remove the effects of exposure on COPD from the model through subtraction, we should also be removing any association between smoking and

exposure in the process. This should leave us with only the association between AN exposure and lung cancer remaining with any confounding by smoking eliminated.

Mathematically, the following is Richardson's method. The parameter we would like to estimate is β_1 which represents the log RR of lung cancer death due to AN exposure adjusted for smoking. We calculate this by computing:

$$\beta_1 = \ln\left(RR_{cancer}^{unadj}\right) - \ln(Bias)$$

where RR_{cancer}^{unadj} represents the RR of lung cancer mortality due to AN exposure and $Bias$ is the bias from confounding. This bias is equal to the RR of smoking given exposure which we estimate as the ratio of those workers exposed to AN who died of COPD to unexposed workers who died of COPD (denoted as RR_{COPD}^{unadj}). The final estimate will then be

$$\hat{\beta}_1 = \ln\left(RR_{cancer}^{unadj}\right) - \ln\left(RR_{COPD}^{unadj}\right).$$

We have assumed that the only confounding effect in the study is due to smoking rather than other factors such as age or genetics. Note that Richardson said in his publication, "To be valid, such an interpretation requires that smoking is related to lung cancer and COPD, there is no true causal association between exposure and COPD, and the only uncontrolled confounder of the association between exposure and COPD is smoking" (2010). The first two conditions are satisfied as mentioned above, but we must assume that the third condition is also true to trust our results.

Step 2: Adjust Relative Risk (RR) Using Estimated Bias

First, risk sets were created for the health outcome (lung cancer death) as described previously. These risk sets were used to perform a conditional logistic regression with AN exposure and smoking history as explanatory variables. The results of this regression analysis are displayed in the fourth column of Table 7. The highlighted row in this table is the adjusted RR using the original smoking data. The estimate of RR = 1.035 (95% CI= (0.97, 1.11) will be used as a comparison for the new estimates.

In the original data set, an insufficient number of COPD cases were available to properly analyze the data as Richardson suggests. To account for the small number of COPD cases, we also used mortality due to heart disease as heart disease has a high risk associated with smoking and is not known to be related to AN exposure. The combination of heart disease mortality and COPD mortality was used to create risk sets and later as the estimate for RR_{COPD}^{unadj}. However, even when including heart disease as an additional cause of death, only 13 cases were observed. Thus, we repeated the analysis, using only cases from all non-malignant respiratory diseases (which includes COPD) to create the risk sets, which provided a total of only nine cases. However, there is some concern that some non-malignant respiratory diseases may be related to AN exposure. These models are shown in Table 8. Each of the rows displays one of the three models mentioned above, and the second column lists the number of cases in each of the situations. A univariate conditional logistic regression was run for each of the three new models, and the results of these are displayed in the fourth column in this table.

Step 3: Analysis of Results

As shown above, using Richardson's method, we calculated *unbiased* RRs using

$$\hat{\beta}_1 = \ln\left(RR_{cancer}^{unadj}\right) - \ln\left(RR_{COPD}^{unadj}\right).$$

The value of RR_{cancer}^{unadj} is contained in the first row in Table 8. This is the result of the conditional logistic regression for lung cancer deaths. The other two rows contain the values of RR_{COPD}^{unadj} (the estimated bias from smoking) which were used in two different applications of Richardson's method.

The $RR_{cancer}^{unadj} = 1.035$ in row one of Table 8 is very close to 1.00 which indicates little relationship between AN exposure and lung cancer. This is consistent with the results found by Marsh et al. (1999). Notice the RR values for COPD and heart disease deaths in column four are very close to 1.00. This indicates one of two things. First, it may show that there is no observable relationship between AN exposure and smoking, as AN exposure did not show an increased risk for COPD and heart disease mortality. Another possible conclusion is that heart disease mortality is not a good cause of death for estimating smoking bias as the relationship between smoking and heart disease mortality is not as strong as the relationship between smoking and COPD mortality. Therefore the bias due to smoking may not be properly represented. Recall, there were too few cases of COPD in the cohort to perform the analysis as Richardson recommends and heart disease deaths were included in an attempt to account for this lack of data. Unfortunately, the use of heart disease mortality in combination with COPD mortality may lead to an incorrect value of RR_{COPD}^{unadj}.

Table 9 displays the results from the computations using Richardson's method. Column 1 lists the different variables considered in the calculations. The label in parentheses indicates which values are used in the calculation of RR_{COPD}^{unadj}. For example, in the first row, the RR for deaths from both COPD and heart disease (from row two of Table 8) is used for RR_{COPD}^{unadj}. In the second column, the lower limit of the confidence interval from deaths by COPD and heart disease is used as the value of RR_{COPD}^{unadj} rather than the point estimate, and the third row uses the value of the upper limit of the confidence interval. The second column in Table 9 displays the adjusted RR which was obtained through the use of Richardson's Method for each of the variables listed in column 1 (exponentiated to return from a logarithmic scale to the original one). The values from column 2 were subtracted from the original $RR_{cancer}^{unadj} = 1.035$ in Table 7, and this difference is shown in column 3 of Table 9. Finally, the 95% confidence interval for the adjusted RR is shown in the last column (computed as instructed in Richardson's paper).

The adjusted RR values $(\hat{\beta}_1)$ displayed in the first three rows of column two of Table 9 are not very different from the RR_{cancer}^{unadj} of 1.035 computed from the original data when smoking history is considered in the analysis. This means that when smoking bias is estimated by the relationship between COPD/heart disease mortality and AN exposure and then removed via Richardson's method, there was very little effect on the RR for lung cancer mortality. This seems to indicate that smoking was not a confounding variable in the relationship between lung cancer death and AN exposure. In the last three rows, however, the RRs computed for respiratory disease deaths differ from 1.00 by a greater amount. When the bias for smoking is accounted for using respiratory disease deaths and removed using Richardson's method, the risk of dying from lung cancer when exposed to acrylonitrile is increased. In this analysis, however, we have assumed that there is no

relationship between AN exposure and respiratory disease mortality which may not be an appropriate assumption.

3.0 DISCUSSION

3.1 COMPARISON OF FIRST AND SECOND METHOD

In this thesis, we analyzed the impact of adjusting for confounding by smoking on the results of the AN exposure study by Marsh et al (1999) using two different methods, Monte Carlo simulation of the smoking data and bias adjustment via Richardson's method. These methods yielded disparate results.

The results of the Monte Carlo simulation revealed what was expected, as we initially believed that smoking habits were positively confounding the RRs. It is well known that smoking is a major risk factor in lung cancer mortality. Thus, smoking should be an important potential confounder when analyzing a relationship between AN exposure and lung cancer. The smoking information in the original cohort was not consistent with this previously known association between smoking and lung cancer as the RR for lung cancer mortality given smoking history was 1.08. A Monte Carlo simulation allowed us to adjust the data to create a more realistic balance of ever/never smokers in the cohort. After running the simulation, we found that the mean RRs of lung cancer death given differing levels of AN exposure decreased after adjusting for the simulated smoking data. Moreover, a larger increase in smoking to lung cancer RRs was correlated with a larger drop in the

exposure vs. lung cancer RRs. In other words, the risk of dying from lung cancer due to exposure to acrylonitrile in the Lima cohort decreased slightly after accounting for positive confounding by smoking in the model.

In Richardson's method, possible confounding by smoking was removed by using Richardson's method. Here, we removed the bias created by a relationship between AN exposure and smoking. Any association between AN exposure and deaths from COPD and heart disease represented a relationship between exposure and smoking history because the two causes are not associated with AN exposure. After running the computations using Richardson's method, the RRs for lung cancer mortality given AN exposure changed very little. This method, then, suggests the association between lung cancer death and AN exposure is not confounded by smoking. In the original Lima cohort, removing the possible bias due to smoking did not change the risk of dying from lung cancer. However, this method was limited by the small number of observed deaths for lung cancer, COPD, and heart disease, and the proper implementation Richardson's method was not possible.

In the Monte Carlo simulation, we concluded that adjusting for confounding by smoking slightly decreases the apparent relationship between dying from lung cancer and exposure to acrylonitrile (even at the highest level of exposure), but in Richardson's method suggested that confounding from smoking was not an issue in the original analysis. However, given this data set and the limited number of cases, the Monte Carlo method provides more meaningful results than the analysis using Richardson's method.

3.2 LIMITATIONS AND STRENGTHS

We conclude that the Monte Carlo simulation method was more meaningful compared to the Richardson method for analyzing the effect of confounding from smoking for this data set. In the Monte Carlo method, we were able to account for the small number of observed deaths by running many simulations and reassigning different smoking histories for the cases and controls. Also, this method is supported by Bayesian theory in that we assumed there was already a relationship between lung cancer and smoking history thereby incorporating a prior distribution within the analysis. When Steenland and Greenland compared the results of MC sensitivity analysis to Bayesian analysis, the results were "similar". For example, they concluded that the "Monte Carlo sensitivity analysis, adjusting for possible confounding by smoking, led to an adjusted standardized mortality ratio of 1.43 (95% Monte Carlo limits: 1.15, 1.78). Bayesian results were similar (95% posterior limits: 1.13, 1.84)." These techniques are further addressed and supported in the paper by Steenland and Greenland (2004).

One issue with the Monte Carlo method is the difficulties in displaying and interpreting the results of the simulations. We calculated 500 different RRs in the simulation, and the correct descriptive statistic must be chosen to aptly convey the appropriate relationship. Here, the mean was chosen, but the median or any of the 500 individual RRs may have been more appropriate for summarizing the results. Additionally, in any simulation, conclusions are not founded on observed results but rather on simulated data. Therefore any interpretation of the results depends on many assumptions made prior to the analysis. In our simulation, we addressed only the most extreme scenario of misclassification in which every lung cancer case except one was reassigned as a smoker. Finally,

the small data set and inadequate number of lung cancer deaths was an issue. This limited the level of simulation possible as there were only 15 cases to reassign.

On the other hand, Richardson's method is much easier to perform, especially if one is not familiar with simulations and the necessary coding. Additionally, the calculations are quick and easy to interpret. However, these seem to be the only advantages of the Richardson's method when applying to our data example. Our analysis was limited, as above, by the small number of lung cancer cases in the original study, so we were unable to run the analysis as Richardson originally recommended. Additionally, there were very few cases of COPD death in the data set. This forced us to include other causes of death to model the association between smoking and AN exposure. These other health issues do not necessarily have the known correlation with smoking and the absence of a relationship with AN exposure.

4.0 CONCLUSIONS

In summary, the Monte Carlo method is more meaningful for analyzing a possible effect of confounding due to smoking in the Lima cohort. In this method, we concluded that smoking did indeed confound the relationship between AN exposure and lung cancer death. After adjusting for the well-known relationship between smoking and lung cancer and the observed relationship between AN exposure and smoking, the RRs for lung cancer death related to acrylonitrile exposure decreased in the Lima cohort, and we observed somewhat less evidence of a relationship between AN exposure and lung cancer.

APPENDIX A: Figure & Tables

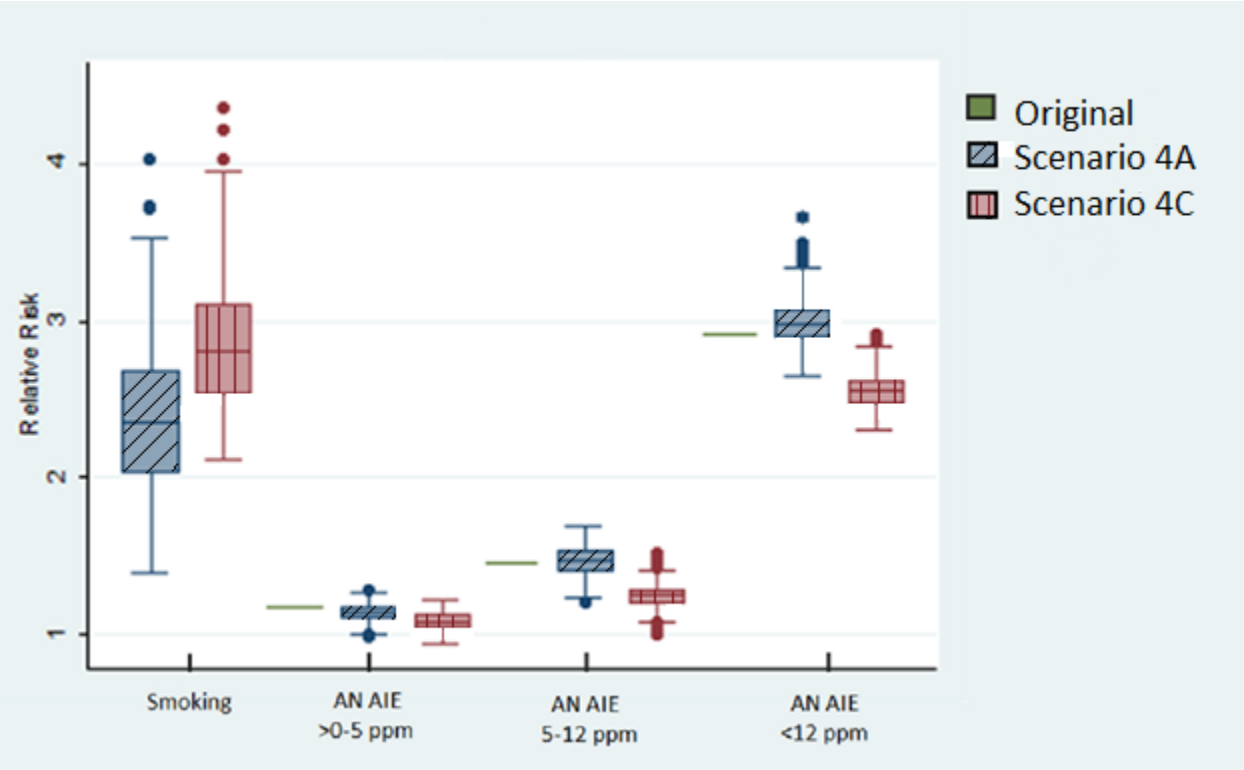

Figure 1: Comparison of Relative Risk Results using Simulated Lima, OH Cohort Data

Table 1: Original Lima, OH Cohort Data-- Lifetime Cigarette Smoking History by AN Exposure [a]

AN_Cum (ppm-years)	Number of Cases	Ever Smoked		Never Smoked		Unknown	
		Number of Obs.	% of Total Cohort	Number of Obs.	% of Total Cohort	Number of Obs.	% of Total Cohort
Unexposed	6	258	54.4	155	32.7	61	12.9
Ever Exposed	9	280	54.0	203	39.2	35	6.8
>0-7.9	2	89	43.2	99	48.1	18	8.7
8.0-109.9	4	130	55.1	91	38.6	15	6.4
110+	3	61	80.3	13	17.1	2	2.6
TOTAL	15	538	54.2	358	36.1	96	9.7

[a] Reported by Marsh et al. (1999)

Table 2: **Original Lima, OH Cohort Data—Summary of Relative Risk Regression Analysis (Univariate Models) for Lung Cancer Mortality** [a]

Variable	Category	Observed Deaths	RR (95% CI)	Global Test p-value
Smoking History	Never Ever Unknown	3 10 2	1.00 1.08 (0.26-6.18) 1.18 (0.09-11.44)	0.999
Duration of AN Exposure	Unexposed >0-4.9 5.0-13.9 14+	6 3 3 3	1.00 1.71 (0.25-8.94) 2.28 (0.35-11.38) 2.15 (0.34-10.70)	0.598
Cumulative AN Exposure	Unexposed >0-7.9 8.0-109.9 110+	6 2 4 3	1.00 1.97 (0.18-12.04) 2.15 (0.43-9.33) 1.97 (0.31-9.42)	0.645
Average Intensity of AN Exposure	Unexposed >0-4.9 5.0-11.9 12.0+	6 3 3 3	1.00 1.97 (0.31-9.54) 1.70 (0.26-8.26) 2.64 (0.42-12.67)	0.513

[a] All RRs are adjusted for age and calendar time via risk set matching

[b] Reported by Marsh et al. (1999)

Table 3: **Original Lima, OH Cohort Data—Summary of Relative Risk Regression Analysis (Bivariate Models [a].) for Lung Cancer Mortality [b.c.]**

Variable	Category	Observed Deaths	RR (95% CI)	Global Test p-value
Duration of AN Exposure	Unexposed >0-4.9 5.0-13.9 14+	6 3 3 3	1.00 1.25 (0.17-7.03) 1.82 (0.26-9.66) 2.20 (0.34-11.24)	0.713
Cumulative AN Exposure	Unexposed >0-7.9 8.0-109.9 110+	6 2 4 3	1.00 1.27 (0.10-8.94) 1.60 (0.29-7.57) 2.19 (0.34-10.70)	0.723
Average Intensity of AN Exposure	Unexposed >0-4.9 5.0-11.9 12.0+	6 3 3 3	1.00 1.18 (0.16-6.84) 1.46 (0.22-7.29) 2.91 (0.46-14.13)	0.514

[a.] Models adjusted for time since first employment (< 20, 20-30, 30+)
[b.] All RRs are adjusted for age and calendar time via risk set matching
[c.] Report by Marsh et al. (1999)

Table 4: Summary of Simulated External Odds Ratios using the Original Lima, OH Cohort

Scenario	Number of Cases Identified as "Ever Smoked"	Number of Cases Identified as "Never Smoked" or "Unknown"	Smoking Prevalence Among Cases	Odds Ratio	Fisher's Exact P-Value
Original	10	5	66.7%	0.985	1.00
1	11	4	73.3%	1.355	0.417
2	12	3	80.0%	1.971	0.408
3	13	2	86.7%	3.202	0.164
4	14	1	93.3%	6.898	0.028*

*Statistically significant at 0.05 level

Table 5: Original Lima, OH Cohort Data—Risk Set Details Used in Monte Carlo Simulations

Case Number	Number of Controls	Ever Smoked Case, Control	Never Smoked Case, Control	Unknown Case, Control
1	22	1, 3	0, 16	0, 3
2	166	0, 40	1, 111	0, 15
3	47	1, 14	0, 30	0, 3
4	91	0, 22	1, 62	0, 7
5	123	1, 27	0, 86	0, 10
6	52	1, 13	0, 34	0, 5
7	114	1, 27	0, 78	0, 9
8	119	1, 30	0, 77	0, 12
9	62	0, 15	1, 41	0, 6
10	78	1, 20	0, 51	0, 7
11	27	1, 5	0, 20	0, 2
12	7	0, 0	0, 4	1, 3
13	79	1, 21	0, 51	0, 7
14	29	1, 7	0, 19	0, 3
15	5	0, 1	0, 4	1, 0
TOTAL	1,021	10, 245	3, 684	2, 92

Table 6: Summary of Conditional Logistic Regression Results using Simulated Lima, OH Cohort Data

Model	Original Univariate [1]			Original Final Model [1,2]		
Summary Statistics	Smoking History	RR	Confidence Interval	AN Exp	RR	Confidence Interval
	Never Ever	1.00 1.08	(0.28-6.18)	Unexposed >0 - 4.9 5.0 -11.9 12.0 +	1.00 1.18 1.46 2.91	 (0.16-6.84) (0.22-7.29) (0.46-14.13)

Model	Scenario 4A [2,3]			Scenario 4B [2,3]			Scenario 4C [2,3]		
Summary Statistics	Category	Mean RR	Std. Dev (Min, Max)	Category	Mean RR	Std. Dev (Min, Max)	Category	Mean RR	Std. Dev (Min, Max)
Smoking History	Never Ever	1.00 1.48	0.43 (0.48, 2.96)	Never Ever	1.00 2.48	0.47 (1.31, 4.34)	Never Ever	1.00 2.92	0.40 (2.04, 4.37
AN Exposure	Unexposed >0 - 4.9 5.0 -11.9 12.0 +	1.00 1.17 1.47 2.99	 0.04 (0.97, 1.24) 0.05 (1.27, 1.69) 0.08 (2.80, 3.39)	Unexposed >0 - 4.9 5.0 -11.9 12.0 +	1.00 1.12 1.37 2.75	 0.06 (0.87, 1.26) 0.07 (1.12, 1.60) 0.14 (2.44, 3.31)	Unexposed >0 - 4.9 5.0 -11.9 12.0 +	1.00 1.07 1.22 2.52	 0.05 (0.91, 1.19) 0.06 (0.99, 1.45) 0.112.30, 2.88)

[1] Reported by Marsh et al. (1999)
[2] Adjusted for time since first employment
[3] Simulated results are based on 500 models

Table 7: Original Lima, OH Cohort Data— Summary of Bivariate Conditional Logistic Regression Results

Outcome Variable	Number of Cases	Independent Variable	RR	95% Confidence Interval
Lung Cancer Death	15	AN Exposure Never Ever	 1.00 1.035	 (0.97,1.11)
		Smoking History Never Ever Unknown	 1.00 1.10 1.18	 (0.30, 4.05) (0.18, 7.73)

Table 8: Original Lima, OH Cohort Data—Summary of Univariate Conditional Logistic Regression Results

Outcome Variable	Number of Cases	Independent Variable	RR	95% Confidence Interval
Lung Cancer Death	15	AN Exposure Never Ever	 1.00 1.034	 (0.96, 1.09)
COPD/ Heart Disease Death	13	AN Exposure Never Ever	 1.00 0.99	 (0.97, 1.03)
Respiratory Disease Deaths	9	AN Exposure Never Ever	 1.00 0.80	 (051, 1.26)

Table 9: Summary of Adjusted Estimated Relative Risks for Lung Cancer Mortality Using Richardson's Method

Bias Variable	$ln(RR_1)$-$ln(RR_2)$	Change in RR	95% Confidence Interval
COPD/HD Deaths (Relative Risk)	1.035	0.0	(0.964, 1.106)
COPD/HD Deaths (Lower Limit of 95% Confidence Interval)	1.07	-0.035	(0.143, 1.497)
COPD/HD Deaths (Upper Limit of 95% Confidence Interval)	1.00	0.035	(0.88, 3.192)
Respiratory Disease Deaths (Relative Risk)	1.292	0.257	(0.136, 2.448)
Respiratory Disease Deaths (Lower Limit of 95% Confidence Interval)	2.03	-0.995	(0.668, 2.480)
Respiratory Disease Deaths (Upper Limit of 95% Confidence Interval)	0.82	0.215	(0.426, 1.238)

APPENDIX B: STATA Code

Scenario 4A – Least Extreme

```
program define neverTOever10
drop smk smk1 smk01 _Ismk1_1 _Itsfe_2 _Itsfe_3 _Iaie_1 _Iaie_2 _Iaie_3
reshape wide  casecontrol aie tsfe, i(  recordnumber) j( casenumber)
generate smk=ever_smk
replace smk=rbinomial(1, 0.5) if smk==9 & ever_case==0
generate smk1=1 if smk==1
replace smk=0 if smk==9
replace smk1=rbinomial(1, 0.67) if smk==0 & ever_case==0 & ever_3==1
replace smk1=rbinomial(1, 0.67) if smk==0 & ever_case==0 & ever_2==1
replace smk1=rbinomial(1, 0.67) if smk==0 & ever_case==0 & ever_1==1
replace smk1=rbinomial(1, 0.67) if smk==0 & ever_case==0 & never==1
reshape long
sort  casecontrol aie tsfe
drop in 1037/5850
*gen smk01=rbinomial(1, 0.67) if smk==0 & ever_case==1
replace smk1=1 if smk1==. & casecon==0
gen smk01= 1 if casecon==1 & smk==1
sort caseco smk01
replace smk01 = 1 in 1036
replace smk01 = 1 in 1035
replace smk01 = 1 in 1034
replace smk01 = 1 in 1033
replace smk01 = 0 in 1032
replace smk1=smk01 if casecon==1
*replace smk1=ever_smk if caseco==0 & ever_smk==0
xi:clog caseco i.smk1 , group(casenumber) or
xi:clog caseco i.smk1 i.aie i.tsfe, group(casenumber) or
end
simulate _b _se, reps(100): neverTOever10



program define neverTOever10
drop smk smk1 smk01 _Ismk1_1 _Itsfe_2 _Itsfe_3 _Iaie_1 _Iaie_2 _Iaie_3
reshape wide  casecontrol aie tsfe, i(  recordnumber) j( casenumber)
generate smk=ever_smk
replace smk=rbinomial(1, 0.5) if smk==9 & ever_case==0
generate smk1=1 if smk==1
replace smk=0 if smk==9
replace smk1=rbinomial(1, 0.67) if smk==0 & ever_case==0 & ever_3==1
replace smk1=rbinomial(1, 0.67) if smk==0 & ever_case==0 & ever_2==1
replace smk1=rbinomial(1, 0.67) if smk==0 & ever_case==0 & ever_1==1
replace smk1=rbinomial(1, 0.67) if smk==0 & ever_case==0 & never==1
reshape long
sort  casecontrol aie tsfe
drop in 1037/5850
*gen smk01=rbinomial(1, 0.67) if smk==0 & ever_case==1
replace smk1=1 if smk1==. & casecon==0
gen smk01= 1 if casecon==1 & smk==1
sort caseco smk01
replace smk01 = 1 in 1036
replace smk01 = 1 in 1035
replace smk01 = 1 in 1034
replace smk01 = 0 in 1033
replace smk01 = 1 in 1032
replace smk1=smk01 if casecon==1
*replace smk1=ever_smk if caseco==0 & ever_smk==0
xi:clog caseco i.smk1 , group(casenumber) or
xi:clog caseco i.smk1 i.aie i.tsfe, group(casenumber) or
end
simulate _b _se, reps(100): neverTOever10



program define neverTOever10
drop smk smk1 smk01 _Ismk1_1 _Itsfe_2 _Itsfe_3 _Iaie_1 _Iaie_2 _Iaie_3
```

```
reshape wide  casecontrol aie tsfe, i(  recordnumber) j( casenumber)
generate smk=ever_smk
replace smk=rbinomial(1, 0.5) if smk==9 & ever_case==0
generate smk1=1 if smk==1
replace smk=0 if smk==9
replace smk1=rbinomial(1, 0.67) if smk==0 & ever_case==0 & ever_3==1
replace smk1=rbinomial(1, 0.67) if smk==0 & ever_case==0 & ever_2==1
replace smk1=rbinomial(1, 0.67) if smk==0 & ever_case==0 & ever_1==1
replace smk1=rbinomial(1, 0.67) if smk==0 & ever_case==0 & never==1
reshape long
sort  casecontrol aie tsfe
drop in 1037/5850
*gen smk01=rbinomial(1, 0.67) if smk==0 & ever_case==1
replace smk1=1 if smk1==. & casecon==0
gen smk01= 1 if casecon==1 & smk==1
sort caseco smk01
replace smk01 = 1 in 1036
replace smk01 = 1 in 1035
replace smk01 = 0 in 1034
replace smk01 = 1 in 1033
replace smk01 = 1 in 1032
replace smk1=smk01 if casecon==1
*replace smk1=ever_smk if caseco==0 & ever_smk==0
xi:clog caseco i.smk1 , group(casenumber) or
xi:clog caseco i.smk1 i.aie i.tsfe, group(casenumber) or
end
simulate _b _se, reps(100): neverTOever10


program define neverTOever10
drop smk smk1 smk01 _Ismk1_1 _Itsfe_2 _Itsfe_3 _Iaie_1 _Iaie_2 _Iaie_3
reshape wide  casecontrol aie tsfe, i(  recordnumber) j( casenumber)
generate smk=ever_smk
replace smk=rbinomial(1, 0.5) if smk==9 & ever_case==0
generate smk1=1 if smk==1
replace smk=0 if smk==9
replace smk1=rbinomial(1, 0.67) if smk==0 & ever_case==0 & ever_3==1
replace smk1=rbinomial(1, 0.67) if smk==0 & ever_case==0 & ever_2==1
replace smk1=rbinomial(1, 0.67) if smk==0 & ever_case==0 & ever_1==1
replace smk1=rbinomial(1, 0.67) if smk==0 & ever_case==0 & never==1
reshape long
sort  casecontrol aie tsfe
drop in 1037/5850
*gen smk01=rbinomial(1, 0.67) if smk==0 & ever_case==1
replace smk1=1 if smk1==. & casecon==0
gen smk01= 1 if casecon==1 & smk==1
sort caseco smk01
replace smk01 = 1 in 1036
replace smk01 = 0 in 1035
replace smk01 = 1 in 1034
replace smk01 = 1 in 1033
replace smk01 = 1 in 1032
replace smk1=smk01 if casecon==1
*replace smk1=ever_smk if caseco==0 & ever_smk==0
xi:clog caseco i.smk1 , group(casenumber) or
xi:clog caseco i.smk1 i.aie i.tsfe, group(casenumber) or
end
simulate _b _se, reps(100): neverTOever10
```

Scenario 4B - Moderate

```
program define neverTOever10
drop smk smk1 smk01 _Ismk1_1 _Itsfe_2 _Itsfe_3 _Iaie_1 _Iaie_2 _Iaie_3
reshape wide  casecontrol aie tsfe, i(  recordnumber) j( casenumber)
generate smk=ever_smk
replace smk=rbinomial(1, 0.5) if smk==9 & ever_case==0
generate smk1=1 if smk==1
replace smk=0 if smk==9
replace smk1=rbinomial(1, 0.8) if smk==0 & ever_case==0 & ever_3==1
replace smk1=rbinomial(1, 0.7) if smk==0 & ever_case==0 & ever_2==1
replace smk1=rbinomial(1, 0.3) if smk==0 & ever_case==0 & ever_1==1
replace smk1=rbinomial(1, 0.3) if smk==0 & ever_case==0 & never==1
reshape long
sort  casecontrol aie tsfe
drop in 1037/5850
*gen smk01=rbinomial(1, 0.67) if smk==0 & ever_case==1
replace smk1=1 if smk1==. & casecon==0
gen smk01= 1 if casecon==1 & smk==1
sort caseco smk01
replace smk01 = 1 in 1036
replace smk01 = 1 in 1035
replace smk01 = 1 in 1034
replace smk01 = 1 in 1033
replace smk01 = 0 in 1032
replace smk1=smk01 if casecon==1
*replace smk1=ever_smk if caseco==0 & ever_smk==0
xi:clog caseco i.smk1 , group(casenumber) or
xi:clog caseco i.smk1 i.aie i.tsfe, group(casenumber) or
end
simulate _b _se, reps(100): neverTOever10



program define neverTOever10
drop smk smk1 smk01 _Ismk1_1 _Itsfe_2 _Itsfe_3 _Iaie_1 _Iaie_2 _Iaie_3
reshape wide  casecontrol aie tsfe, i(  recordnumber) j( casenumber)
generate smk=ever_smk
replace smk=rbinomial(1, 0.5) if smk==9 & ever_case==0
generate smk1=1 if smk==1
replace smk=0 if smk==9
replace smk1=rbinomial(1, 0.8) if smk==0 & ever_case==0 & ever_3==1
replace smk1=rbinomial(1, 0.7) if smk==0 & ever_case==0 & ever_2==1
replace smk1=rbinomial(1, 0.3) if smk==0 & ever_case==0 & ever_1==1
replace smk1=rbinomial(1, 0.3) if smk==0 & ever_case==0 & never==1
reshape long
sort  casecontrol aie tsfe
drop in 1037/5850
*gen smk01=rbinomial(1, 0.67) if smk==0 & ever_case==1
replace smk1=1 if smk1==. & casecon==0
gen smk01= 1 if casecon==1 & smk==1
sort caseco smk01
replace smk01 = 1 in 1036
replace smk01 = 1 in 1035
replace smk01 = 1 in 1034
replace smk01 = 0 in 1033
replace smk01 = 1 in 1032
replace smk1=smk01 if casecon==1
*replace smk1=ever_smk if caseco==0 & ever_smk==0
xi:clog caseco i.smk1 , group(casenumber) or
xi:clog caseco i.smk1 i.aie i.tsfe, group(casenumber) or
end
simulate _b _se, reps(100): neverTOever10



program define neverTOever10
drop smk smk1 smk01 _Ismk1_1 _Itsfe_2 _Itsfe_3 _Iaie_1 _Iaie_2 _Iaie_3
reshape wide  casecontrol aie tsfe, i(  recordnumber) j( casenumber)
generate smk=ever_smk
replace smk=rbinomial(1, 0.5) if smk==9 & ever_case==0
generate smk1=1 if smk==1
```

```
replace smk=0 if smk==9
replace smk1=rbinomial(1, 0.8) if smk==0 & ever_case==0 & ever_3==1
replace smk1=rbinomial(1, 0.7) if smk==0 & ever_case==0 & ever_2==1
replace smk1=rbinomial(1, 0.3) if smk==0 & ever_case==0 & ever_1==1
replace smk1=rbinomial(1, 0.3) if smk==0 & ever_case==0 & never==1
reshape long
sort  casecontrol aie tsfe
drop in 1037/5850
*gen smk01=rbinomial(1, 0.67) if smk==0 & ever_case==1
replace smk1=1 if smk1==. & casecon==0
gen smk01= 1 if casecon==1 & smk==1
sort caseco smk01
replace smk01 = 1 in 1036
replace smk01 = 1 in 1035
replace smk01 = 0 in 1034
replace smk01 = 1 in 1033
replace smk01 = 1 in 1032
replace smk1=smk01 if casecon==1
*replace smk1=ever_smk if caseco==0 & ever_smk==0
xi:clog caseco i.smk1 , group(casenumber) or
xi:clog caseco i.smk1 i.aie i.tsfe, group(casenumber) or
end
simulate _b _se, reps(100): neverTOever10



program define neverTOever10
drop smk smk1 smk01 _Ismk1_1 _Itsfe_2 _Itsfe_3 _Iaie_1 _Iaie_2 _Iaie_3
reshape wide  casecontrol aie tsfe, i(  recordnumber) j( casenumber)
generate smk=ever_smk
replace smk=rbinomial(1, 0.5) if smk==9 & ever_case==0
generate smk1=1 if smk==1
replace smk=0 if smk==9
replace smk1=rbinomial(1, 0.8) if smk==0 & ever_case==0 & ever_3==1
replace smk1=rbinomial(1, 0.7) if smk==0 & ever_case==0 & ever_2==1
replace smk1=rbinomial(1, 0.3) if smk==0 & ever_case==0 & ever_1==1
replace smk1=rbinomial(1, 0.3) if smk==0 & ever_case==0 & never==1
reshape long
sort  casecontrol aie tsfe
drop in 1037/5850
*gen smk01=rbinomial(1, 0.67) if smk==0 & ever_case==1
replace smk1=1 if smk1==. & casecon==0
gen smk01= 1 if casecon==1 & smk==1
sort caseco smk01
replace smk01 = 1 in 1036
replace smk01 = 0 in 1035
replace smk01 = 1 in 1034
replace smk01 = 1 in 1033
replace smk01 = 1 in 1032
replace smk1=smk01 if casecon==1
*replace smk1=ever_smk if caseco==0 & ever_smk==0
xi:clog caseco i.smk1 , group(casenumber) or
xi:clog caseco i.smk1 i.aie i.tsfe, group(casenumber) or
end
simulate _b _se, reps(100): neverTOever10



program define neverTOever10
drop smk smk1 smk01 _Ismk1_1 _Itsfe_2 _Itsfe_3 _Iaie_1 _Iaie_2 _Iaie_3
reshape wide  casecontrol aie tsfe, i(  recordnumber) j( casenumber)
generate smk=ever_smk
replace smk=rbinomial(1, 0.5) if smk==9 & ever_case==0
generate smk1=1 if smk==1
replace smk=0 if smk==9
replace smk1=rbinomial(1, 0.8) if smk==0 & ever_case==0 & ever_3==1
replace smk1=rbinomial(1, 0.7) if smk==0 & ever_case==0 & ever_2==1
replace smk1=rbinomial(1, 0.3) if smk==0 & ever_case==0 & ever_1==1
replace smk1=rbinomial(1, 0.3) if smk==0 & ever_case==0 & never==1
reshape long
sort  casecontrol aie tsfe
drop in 1037/5850
```

```
*gen smk01=rbinomial(1, 0.67) if smk==0 & ever_case==1
replace smk1=1 if smk1==. & casecon==0
gen smk01= 1 if casecon==1 & smk==1
sort caseco smk01
replace smk01 = 0 in 1036
replace smk01 = 1 in 1035
replace smk01 = 1 in 1034
replace smk01 = 1 in 1033
replace smk01 = 1 in 1032
replace smk1=smk01 if casecon==1
*replace smk1=ever_smk if caseco==0 & ever_smk==0
xi:clog caseco i.smk1 , group(casenumber) or
xi:clog caseco i.smk1 i.aie i.tsfe, group(casenumber) or
end
simulate _b _se, reps(100): neverTOever10
```

Scenario 4C – Most Extreme

```
program define neverTOever10
drop smk smk1 smk01 _Ismk1_1 _Itsfe_2 _Itsfe_3 _Iaie_1 _Iaie_2 _Iaie_3
reshape wide  casecontrol aie tsfe, i(  recordnumber) j( casenumber)
generate smk=ever_smk
replace smk=rbinomial(1, 0.5) if smk==9 & ever_case==0
generate smk1=1 if smk==1
replace smk=0 if smk==9
replace smk1=rbinomial(1, 0.95) if smk==0 & ever_case==0 & ever_3==1
replace smk1=rbinomial(1, 0.9) if smk==0 & ever_case==0 & ever_2==1
replace smk1=rbinomial(1, 0.1) if smk==0 & ever_case==0 & ever_1==1
replace smk1=rbinomial(1, 0.05) if smk==0 & ever_case==0 & never==1
reshape long
sort  casecontrol aie tsfe
drop in 1037/5850
*gen smk01=rbinomial(1, 0.67) if smk==0 & ever_case==1
replace smk1=1 if smk1==. & casecon==0
gen smk01= 1 if casecon==1 & smk==1
sort caseco smk01
replace smk01 = 1 in 1036
replace smk01 = 1 in 1035
replace smk01 = 1 in 1034
replace smk01 = 1 in 1033
replace smk01 = 0 in 1032
replace smk1=smk01 if casecon==1
*replace smk1=ever_smk if caseco==0 & ever_smk==0
xi:clog caseco i.smk1 , group(casenumber) or
xi:clog caseco i.smk1 i.aie i.tsfe, group(casenumber) or
end
simulate _b _se, reps(100): neverTOever10



program define neverTOever10
drop smk smk1 smk01 _Ismk1_1 _Itsfe_2 _Itsfe_3 _Iaie_1 _Iaie_2 _Iaie_3
reshape wide  casecontrol aie tsfe, i(  recordnumber) j( casenumber)
generate smk=ever_smk
replace smk=rbinomial(1, 0.5) if smk==9 & ever_case==0
generate smk1=1 if smk==1
replace smk=0 if smk==9
replace smk1=rbinomial(1, 0.95) if smk==0 & ever_case==0 & ever_3==1
replace smk1=rbinomial(1, 0.9) if smk==0 & ever_case==0 & ever_2==1
replace smk1=rbinomial(1, 0.1) if smk==0 & ever_case==0 & ever_1==1
replace smk1=rbinomial(1, 0.05) if smk==0 & ever_case==0 & never==1
reshape long
sort  casecontrol aie tsfe
drop in 1037/5850
*gen smk01=rbinomial(1, 0.67) if smk==0 & ever_case==1
replace smk1=1 if smk1==. & casecon==0
gen smk01= 1 if casecon==1 & smk==1
sort caseco smk01
replace smk01 = 1 in 1036
replace smk01 = 1 in 1035
replace smk01 = 1 in 1034
replace smk01 = 0 in 1033
replace smk01 = 1 in 1032
replace smk1=smk01 if casecon==1
*replace smk1=ever_smk if caseco==0 & ever_smk==0
xi:clog caseco i.smk1 , group(casenumber) or
xi:clog caseco i.smk1 i.aie i.tsfe, group(casenumber) or
end
simulate _b _se, reps(100): neverTOever10



program define neverTOever10
drop smk smk1 smk01 _Ismk1_1 _Itsfe_2 _Itsfe_3 _Iaie_1 _Iaie_2 _Iaie_3
reshape wide  casecontrol aie tsfe, i(  recordnumber) j( casenumber)
generate smk=ever_smk
replace smk=rbinomial(1, 0.5) if smk==9 & ever_case==0
generate smk1=1 if smk==1
```

```
replace smk=0 if smk==9
replace smk1=rbinomial(1, 0.95) if smk==0 & ever_case==0 & ever_3==1
replace smk1=rbinomial(1, 0.9) if smk==0 & ever_case==0 & ever_2==1
replace smk1=rbinomial(1, 0.1) if smk==0 & ever_case==0 & ever_1==1
replace smk1=rbinomial(1, 0.05) if smk==0 & ever_case==0 & never==1
reshape long
sort  casecontrol aie tsfe
drop in 1037/5850
*gen smk01=rbinomial(1, 0.67) if smk==0 & ever_case==1
replace smk1=1 if smk1==. & casecon==0
gen smk01= 1 if casecon==1 & smk==1
sort caseco smk01
replace smk01 = 1 in 1036
replace smk01 = 1 in 1035
replace smk01 = 0 in 1034
replace smk01 = 1 in 1033
replace smk01 = 1 in 1032
replace smk1=smk01 if casecon==1
*replace smk1=ever_smk if caseco==0 & ever_smk==0
xi:clog caseco i.smk1 , group(casenumber) or
xi:clog caseco i.smk1 i.aie i.tsfe, group(casenumber) or
end
simulate _b _se, reps(100): neverTOever10



program define neverTOever10
drop smk smk1 smk01 _Ismk1_1 _Itsfe_2 _Itsfe_3 _Iaie_1 _Iaie_2 _Iaie_3
reshape wide  casecontrol aie tsfe, i(  recordnumber) j( casenumber)
generate smk=ever_smk
replace smk=rbinomial(1, 0.5) if smk==9 & ever_case==0
generate smk1=1 if smk==1
replace smk=0 if smk==9
replace smk1=rbinomial(1, 0.95) if smk==0 & ever_case==0 & ever_3==1
replace smk1=rbinomial(1, 0.9) if smk==0 & ever_case==0 & ever_2==1
replace smk1=rbinomial(1, 0.1) if smk==0 & ever_case==0 & ever_1==1
replace smk1=rbinomial(1, 0.05) if smk==0 & ever_case==0 & never==1
reshape long
sort  casecontrol aie tsfe
drop in 1037/5850
*gen smk01=rbinomial(1, 0.67) if smk==0 & ever_case==1
replace smk1=1 if smk1==. & casecon==0
gen smk01= 1 if casecon==1 & smk==1
sort caseco smk01
replace smk01 = 1 in 1036
replace smk01 = 0 in 1035
replace smk01 = 1 in 1034
replace smk01 = 1 in 1033
replace smk01 = 1 in 1032
replace smk1=smk01 if casecon==1
*replace smk1=ever_smk if caseco==0 & ever_smk==0
xi:clog caseco i.smk1 , group(casenumber) or
xi:clog caseco i.smk1 i.aie i.tsfe, group(casenumber) or
end
simulate _b _se, reps(100): neverTOever10



program define neverTOever10
drop smk smk1 smk01 _Ismk1_1 _Itsfe_2 _Itsfe_3 _Iaie_1 _Iaie_2 _Iaie_3
reshape wide  casecontrol aie tsfe, i(  recordnumber) j( casenumber)
generate smk=ever_smk
replace smk=rbinomial(1, 0.5) if smk==9 & ever_case==0
generate smk1=1 if smk==1
replace smk=0 if smk==9
replace smk1=rbinomial(1, 0.95) if smk==0 & ever_case==0 & ever_3==1
replace smk1=rbinomial(1, 0.9) if smk==0 & ever_case==0 & ever_2==1
replace smk1=rbinomial(1, 0.1) if smk==0 & ever_case==0 & ever_1==1
replace smk1=rbinomial(1, 0.05) if smk==0 & ever_case==0 & never==1
reshape long
sort  casecontrol aie tsfe
drop in 1037/5850
```

```
*gen smk01=rbinomial(1, 0.67) if smk==0 & ever_case==1
replace smk1=1 if smk1==. & casecon==0
gen smk01= 1 if casecon==1 & smk==1
sort caseco smk01
replace smk01 = 0 in 1036
replace smk01 = 1 in 1035
replace smk01 = 1 in 1034
replace smk01 = 1 in 1033
replace smk01 = 1 in 1032
replace smk1=smk01 if casecon==1
*replace smk1=ever_smk if caseco==0 & ever_smk==0
xi:clog caseco i.smk1 , group(casenumber) or
xi:clog caseco i.smk1 i.aie i.tsfe, group(casenumber) or
end
simulate _b _se, reps(100): neverTOever10
```

BIBLIOGRAPHY

Benn T, Osborne K. Mortality of United Kingdom acrylonitrile workers: an extended and updated study. Scand J Work Environ Health 1998; 24 suppl 2:17-24.

Blair A, Stewart PA, Zaebest D, et al. Mortality study of industrial workers exposed to acrylonitrile. Scand J Work Environ Health 1998; 24 Suppl 2:25-41.

Cunningham M. Reevaluation of lung cancer risk in acrylonitrile workers by simulation of missing smoking status information. Unpublished MS Thesis, University of Pittsburgh, Department of Biostatistics: 2005.

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 of Empid 2004; 160:384-392.

Higgins IT, Wynder EL. Reduction in risk of lung cancer among ex-smokers with particular reference to histologic type. Cancer 1988; 62(11):2397-2401.

International Agency for Research on Cancer (IARC). Acrylonitrile, part 1. Lyon: France, 1999: 43-108. IARC monographs on the evaluation of the carcinogenic risks to humans, vol. 71.

Loeb LA, Ernster VL, et al. Smoking and lung cancer: an overview. Cancer Research 1984; 44:5940-5958.

Marsh GM, Sachs D, et al. Direct methods of obtaining information on cigarette smoking in occupational studies. Am J Ind Med 1988; 13:71-103.

Marsh GM, Youk AO, Stone R et al. OCMAP-PLUS: a program for the comprehensive analysis of occupational cohort data. J Occup Environ Med 1998; 40:351-62.

Marsh GM, Gula MJ, Youk AO, Schall LC. Mortality among chemical plant workers exposed to acrylonitrile and other substances. Am J Ind Med 1999; 36:423-436.

Richardson, DB. Occupational exposures and lung cancer adjustment for unmeasured confounding by smoking. Epidemiology 2010; Vol 21 No 2: 181-186.

Samet, JM. The Epidemiology of lung cancer. Chest Journal 1993; 103 Suppl 20S-29S.

StataCorp. 2005. Stata Statistical Software: Release 9. College Station, TX: StataCorp LP.

Strother DE, Mast RW, Kraska RC, Frankos V. Acrylonitrile as a carcinogen: research needs for better risk assessment. Ann NY Acad Sci 1988; 534: 169-78.

Swaen GMH, Bloemen LJN, Twisk J, et al. Mortality update of workers exposed to acrylonitrile in The Netherlands. Scand J Work Environ Health 1998 24 suppl 2:10-16.

Symons JM, Kreckmann KH, et al. Mortality among workers exposed to acrylonitrile in fiber production: an update. J Occup Environ Med 2008; 50(5):550-560.

U.S. Dept. of Health and Human Services, Public Health Service, Office of the Surgeon General. The health consequences of smoking: cancer and chronic lung disease in the workplace: A Report of the Surgeon General. Rockville, MD: DHSS, 1986.

U.S. Dept. of Health and Human Services, Public Health Service, Office of the Surgeon General. How tobacco smoke causes disease: the biology and behavioral basis for smoking-attributable disease; A Report of the Surgeon General. Rockville, MD: DHSS, 2010.

US Evironmental Protection Agency. Integrated risk information system (IRIS) on Acrylonitrile. Cincinnati (OH): Environmental Criteria and Assessment Office, Office of Health and Environment Assessments, Office of Research and Development, 1993.