1st Solicitation for Single Investigator Research Grants

(AFC113)

ALPHA FOUNDATION FOR THE IMPROVEMENT OF MINE SAFETY AND HEALTH

Final Technical Report

1.0 Cover Page

Grant Title: Ischemic Heart Disease and Lung Cancer Mortality in Relation to Respirable Particulate Matter and Diesel Exhaust in Non-metal Miners

Grant Number: AFC113-08

Grant Title: Ischemic Heart Disease and Lung Cancer Mortality in Relation to Respirable Particulate Matter and Diesel Exhaust in Non-metal Miners

Organization: The Regents of the University of California (University of California Berkeley)

Principle Investigator: Ellen A. Eisen, ScD

Contact information: eeisen@berkeley.edu
PH: (510) 643-5310
Fax: (510) 642-5815

Period of Performance: November 1, 2013 – January 31, 2017

Acknowledgement/Disclaimer:

This study was sponsored by the Alpha Foundation for the Improvement of Mine Safety and Health, Inc. (ALPHA FOUNDATION). The views, opinions, and recommendations expressed, as well as the mention of any company name, product or software, herein are solely those of the authors and do not imply any endorsement by the ALPHA FOUNDATION, its Directors and staff.

Section 2.0 Executive Summary

Miners are exposed to far higher levels of diesel exhaust and respirable particulate matter (RPM) than are found in urban ambient environments or other industrial settings in the United States. PM in traffic-related air pollution is recognized as an important risk factor for ischemic heart disease (IHD) based on a vast epidemiologic literature, yet heart disease due to RPM or respirable elemental carbon (REC) from diesel exhaust has rarely been studied in working populations. Diesel exhaust has been associated with lung cancer in miners and RPM exposure with coal workers' pneumoconiosis among relatively young miners. However, if either of these exposures also contributes to the risk of heart disease in non-coal miners, then the total disease burden for miners would be far greater.

We took advantage of the Diesel Exhaust in Miners Study (DEMS), jointly conducted by NCI and NIOSH to study lung cancer and diesel exhaust in non-metal miners, and to examine IHD mortality in relation to diesel exhaust and RPM. The original DEMS investigators published a cohort and a nested case-control study providing evidence of increased risk of lung cancer with increased diesel exhaust exposure. Additionally, they reported no evidence that the DEMS miners were at increased risk for IHD, based on a summary contrast with the general population, SMR = 0.99 (95% CI: 0.91, 1.07). Our project was motivated by the suspicion that the null results for IHD, as well as the positive results for lung cancer, were both likely attenuated by the healthy worker effect. Our goal was to conduct an internal analysis of the miners in DEMS to examine the risk of IHD mortality, as well as lung cancer, in relation to respirable diesel exhaust and PM exposure, paying special attention to the shape of the exposure-response curve, the timing of exposure, and control of healthy worker bias.

Results of our study provide evidence that exposure to diesel exhaust and RPM may increase risk of IHD mortality among workers in the DEMS cohort. We used two approaches to examine the question, the standard method and a more sophisticated g-method to address healthy worker survivor bias. In the standard analysis, the hazard ratios (HRs) for the highest exposure category versus the lowest were 1.18 (95% CI: 0.56, 2.24) for cumulative (REC), 1.25 (95% CI: 0.78, 2.01) for cumulative RPM, 0.75 (95% CI: 0.39, 1.44) for average intensity REC, and 2.58 (95% CI: 1.26, 5.28) for average intensity RPM. The timing of exposure mattered in that the HRs were higher during active employment when exposure was still occurring up to the time of death. The HR for IHD and cumulative RPM (mg/m^3-years) was stronger among those who died while still actively employed (HR= 1.015 per mg/m^3-year) than for all workers in the cohort (HR= 1.004 per mg/m^3-yr). Exposure-response models were stratified by location (surface only and ever-underground). Among ever-underground workers, both diesel exhaust and RPM were associated with elevated rates of IHD mortality; among surface-only workers only RPM was associated with IHD.

In a companion analysis, we applied a method called parametric g-formula that provides results that account for the possible self-selection of less healthy workers out of the workforce. Using this approach, we estimated the risk under a hypothetical intervention where annual average daily exposures to REC is set to 0 and a joint intervention consistent with REC and RPM exposure limits of 0 and 0.5 mg/m^3 respectively. The ratio comparing the risk under the intervention on REC alone and for the joint intervention, each compared to the observed risk, was 0.86 (0.62, 1.17) and 0.84 (0.71, 0.98) respectively. The excess fraction of IHD deaths due to the observed diesel exhaust and RPM exposures in this population was estimated to be 17%.

In our reanalysis of lung cancer in the DEMS population, we compared the lifetime probability of lung cancer among ever-underground workers under the same exposure limits for REC described above with results under no intervention (natural course, i.e., what actually happened). The lifetime cumulative incidence when everyone was always unexposed to REC was estimated at 5.7%, for a risk difference of 2.1%, compared with the observed risk of 7.8%. The attributable fraction of lung cancer deaths due to the observed diesel exhaust exposures in this population was estimated at 27%.

Risk estimates comparing IHD mortality under hypothetical interventions on REC and RPM exposure what actually happened, were smaller in magnitude than our estimates of the impact from the same hypothetical exposure interventions on lung cancer. IHD, however, is a much more common cause of death than lung cancer. Therefore, the risk reduction associated with exposure interventions was greater for IHD than for lung cancer on the absolute scale, i.e., the number of deaths, which is a more meaningful measure of public health impact. The number of IHD deaths prevented due to a reduction in exposure would be greater than the number of lung cancer deaths prevented by the same intervention, despite the apparent weaker exposure-response relationship on the relative scale.

3.0 Problem Statement and Objective

Focus area: Health, Cardiovascular Disease

3.1 Problem Statement

Miners are exposed to far higher levels of respirable particulate matter (RPM) and diesel exhaust than are found in urban ambient environments in the United States. The US Environmental Protection Agency standard for Particulate Matter $(PM)_{2.5}$ is 0.035 mg/m^3 – two orders of magnitude lower than the Mining Safety and Health Administration (MSHA) standard for RPM (currently set at 1.5 mg/m^3).[1] Yet while PM in traffic-related air pollution is recognized as an important risk factor for ischemic heart disease (IHD) based on a vast epidemiologic literature,[2–4] heart disease has rarely been studied in working populations.[5]

One study that did examine the relationship between diesel exhaust and IHD among non-metal miners, Attfield et al, reported that the workers in the Diesel Exhaust in Miners Study (DEMS) were not at increased risk for IHD compared to the general population (SMR = 0.99, 95% CI (0.91, 1.07)).[6] This finding was likely attenuated by both the healthy hire effect and the healthy worker survivor effect (HWSE), which are the two aspects of the healthy worker effect. HWSE occurs in longitudinal studies when workers reduce their exposures for health-related reasons, leading to more accumulated exposure for the least susceptible workers, and ultimately resulting in downward bias that is harder to eliminate. With respect to lung cancer, Attfield et al found that the exposure-response for diesel exhaust was not as steep for underground miners as it was for surface miners[6] – a counterintuitive result that might be explained by a stronger HWSE among miners working underground.

Obtaining valid and non-biased estimates for the exposure-response relations between RPM, diesel exhaust, and IHD mortality is crucial for guiding future MSHA regulations in non-metal mines. In mid-2016, MSHA lowered the concentration limits for RPM exposure from 2 mg/m^3 to 1.5 mg/m^3 based on the rising rates of black lung diseases (including pneumoconiosis) among relatively young coal miners.[1] This does not apply to the RPM limit in the non-coal mines, where the *mean* concentration of RPM underground is close to 1.5 mg/m^3.

3.2 Objectives

Our research attempted to solve the problem of estimating the exposure-response relations between RPM, diesel exhaust, and both lung cancer and IHD mortality in a cohort of miners without bias due to the HWSE or confounding by cigarette smoking. DEMS, a study originally designed to examine lung cancer, provided us the opportunity to examine IHD mortality in relation to both RPM (RPM) and diesel exhaust, measured as respirable elemental carbon (REC), with a focus on bias reduction and causal inference. We have also applied the same focus on bias reduction to a reanalysis of lung cancer mortality in the DEMS cohort in order to provide exposure effect estimates unbiased by the healthy worker survivor effect.

3.3 Specific Aims

Aim 1: Estimate the exposure-response for IHD mortality in relation to cumulative exposure to RPM and REC in the DEMS cohort using Cox proportional hazard models with penalized splines to allow non-linearity.

Aim 2: Distinguish effects of current (or recent) exposure from cumulative exposure on IHD mortality by looking at time windows of exposure among the actively employed miners.

Aim 3: Correct the exposure-response analyses for bias due to the healthy worker survivor effect.

3a: Adjust for bias from *left truncation*, the phenomenon of healthier workers remaining at work until the start of follow up, by restricting the analysis to workers with a later hire date.

3b: Adjust for bias from *right truncation*, the phenomenon of healthier workers staying at work through the end of follow up, by incorporating censoring weights in a Cox model.

3c: Use g-estimation methods to incorporate information on intermittent time off work and leaving work, time-varying indicators of health status likely to be on the pathway from exposure to disease.

Aim 4: Indirectly adjust for potential confounding by smoking by using covariate data collected for a case-control study of lung cancer nested in the cohort.

Aim 5: Refine the exposure-response associations recently reported between REC and lung cancer mortality in surface and underground non-metal miners adjusting for the HWSE and smoking.

4.0 Research Approach

Our overall research approach was to use exposure, demographic and outcome data from an existing cohort mortality study of 12,315 non-coal, non-metal miners to conduct epidemiologic analyses to estimate the exposure-response relationships between RPM, diesel exhaust, and both IHD and lung cancer mortality in a cohort of miners without bias due to HWSE or confounding by cigarette smoking. In the following section, we detail our research approach for data acquisition and analytic methods.

4.1 Data Acquisition

Three distinct but linkable data files were required for our research:

1) The original DEMS cohort data with lung cancer outcomes
2) The nested case-control study with data on cigarette smoking
3) IHD mortality data to link to the cohort

We submitted requests to the appropriate institutions in order to receive these files (National Institute for Occupational Safety and Health (NIOSH), the National Cancer Institute (NCI), and the National Center for Health Statistics (NCHS). Once we received access to the files, we linked them together by subject identifier and subsequently completed our statistical analyses to estimate the unbiased exposure-response relationships between RPM, diesel exhaust, and both IHD and lung cancer mortality.

We experienced significant delays in acquiring the data. These delays were described in detail in Progress Reports. The data were collected, analyzed and published jointly by investigators at NIOSH and NCI. During our grant period, NIOSH was in the process of turning over the stewardship of the data to NCI due to the ongoing resource demands. This transition was responsible for some of the delays in our getting the access we needed.

Ultimately we were successful in obtaining all of the necessary data by developing a strong collaborative relationship with the two principal investigators of DEMS – Dr. Michael Attfield (NIOSH) and Dr. Debra Silverman (NCI). Drs Silverman and Attfield, as well as exposure assessment colleagues Trish Stewart (NIOSH) and Roel Vermeulen (NCI), epidemiologist Aaron Blair (NCI), and statistician Jay Lubin (NCI), are all co-authors on the two IHD papers currently in draft (See two companion papers in preparation).

4.2 IHD Mortality Analysis

Given our overall research goal of estimating the exposure-response relationships between RPM, diesel exhaust, and IHD mortality in a cohort of miners without bias due to HWSE or confounding by cigarette smoking, we moved forward with our statistical analyses using a four-step approach.

4.2.1 Estimate Exposure-Response using Cox Models with Splines

The shape of the exposure response curve for particulates and heart disease is unknown, and the exposure-response relationships of interest may be non-linear in the DEMS cohort. Categorical models are often used to describe nonlinear exposure-response relationships, however these models assume uniform risk within categories and the selection of cut-points can influence the observed results.[7] In light of these disadvantages, we first estimated the exposure-response relationships for cumulative exposure to RPM and REC and IHD mortality in the DEMS cohort using Cox proportional hazard models with penalized splines to allow non-linearity.[8]

4.2.2 Examine the Effects of Different Windows of Exposure

Like many occupational cohorts, the DEMS cohort is designed to study exposures cumulated over a working life in relation to outcomes that occur well after the cessation of employment. However, the relevant time window for heart disease may be more proximate to the event. In fact, most literature on PM-related IHD risk are ambient air pollution studies of short term exposure.[9-11] Cigarette smoking, another major risk factor for IHD, also has a short-term impact; the risk of IHD drops dramatically for ex-smokers within the first year after cessation of cigarette smoking.[12-14]

Therefore, we performed analyses to distinguish the effects of recent exposure from cumulative exposure on IHD mortality. Because only active employees can be currently exposed in an occupational setting, we restricted our analysis of the impact of recent exposure on IHD mortality to actively employed (within 6 months) person time.

4.2.3 Healthy Worker Effect Bias Reduction

Results of many occupational epidemiology studies of chronic diseases are likely attenuated by the healthy worker effect. The healthy worker effect consists of two parts: (1) the healthy hire effect, in which people who are hired into jobs are healthier than those who are not, and (2) the healthy worker survivor effect (HWSE), in which workers reduce their exposures by transferring to a lower exposure job, taking time off work, or leaving employment altogether for health-related reasons. This leads to more accumulated exposure among the healthiest workers.

We have eliminated the healthy hire effect by performing internal analysis, comparing more-exposed to less-exposed workers, rather than comparing workers to the general population.

Bias from HWSE is more difficult to eliminate. The HWSE can arise in several ways depending on the nature of the study design: at the start of follow up (left truncation), at the end of follow up (right truncation or censoring), or throughout the study period in the form of job transfer or temporary time off work during employment.

Attenuation in the exposure-response relationship from left-truncation bias occurs if those who left work prior to the start of follow up were more susceptible to the health effects of exposure than those who

remained to be included in the study.[15] Restricting occupational cohorts to include only workers hired closer to the start of follow-up can reduce left-truncation bias. In DEMS, follow-up started in the year of dieselization in each of the eight mines (1947 to 1967) and continued through 1997. Although there was no exposure to diesel prior to start of follow-up, left truncation is a potential bias in DEMS and needs to be addressed because miners who were at work in the year of dieselization had remained employed despite other sources of RPM.

Attenuation in the exposure-response relationship from right-censoring may occur if individuals who leave work, and thus are no longer exposed, are at greater risk of the adverse health outcome. With work status as a time-varying confounder, standard statistical methods are not adequate to estimate an unbiased effect of the exposure on the outcome when work status is also affected by prior exposure. Thus, if termination of employment is affected by exposure, the effect of exposure on IHD mortality may be even higher than previously reported. To account for right-censoring, we evaluated the relationship between REC exposure and time to termination of employment using an accelerated failure time model. We then used the parametric g-formula to assess the impacts of hypothetical interventions on REC exposure on the risk of IHD mortality, controlling for work status. The parametric g-formula is one of a broader group of "g-methods"[16] first introduced by Robins[17, 18] to control for time-varying confounding affected by past exposure.

By implementing these approaches, we have been able to correct the exposure-response analyses for bias due to the HWSE.

4.2.4 Adjustment for Smoking

We leveraged the information collected on smoking in the nested case-control study to adjust for smoking in our analyses of IHD. To do this, we used smoking data available on the smaller number of subjects in the nested case-control study to impute smoking status in the full cohort using two different multiple imputation methods.

4.3 Lung Cancer Mortality Analysis

Utilizing the same methods as described above in section 4.2.3 (Healthy Worker Effect Bias Reduction), we employed methods to adjust for HWSE in our re-analysis of the exposure-response relationship between REC and lung cancer mortality. Specifically, we evaluated the relationship between REC exposure and time to termination of employment using an accelerated failure time model. We then used the parametric g-formula to assess the impacts of hypothetical interventions on REC exposure on the risk of lung cancer mortality, controlling for work status, thereby controlling for time-varying confounding affected by past exposure.

5.0 Summary of Accomplishments

5.1 Data Acquisition

Originally, we had planned on receiving a file with cause of death data (International Classification of Diseases (ICD) codes) linked to each of the DEMS cohort subjects, but the NIOSH/NCI agreement with the National Death Index had lapsed and we were unable to get the data through this route. Instead, we submitted a proposal to the National Center for Health Statistics (NCHS) to access cause of death data for DEMS cohort subjects.

Applying to NCHS for, and accessing, restricted data is complex and is run through NCHS's Research Data Center (RDC). In order to be approved for accessing these data, our researchers were required to undergo special training to obtain "Special Sworn Status" with the United States Census Bureau. We were also required to go through a complete background check to obtain the appropriate level of security clearance.

The modes for accessing these restricted data through NCHS are very limited. One mode of access is through the network of Federal Statistical Research Data Centers (FSRDCs), which is managed by the Census Bureau. In 2016, we were approved to access data following standard RDC procedures for accessing non-NCHS restricted data using the California RDC (part of the FSRDC network), which is located on the UC Berkeley campus.

The California RDC (and other FSRDCs) are physical computer terminals through which all data must be accessed. No outside data may be brought in; instead researchers are given a folder containing all of their data, and programs must be run on-site at the FSRDC through an RDC server. Although we had already obtained access to the DEMS cohort data and the nested case-control data, we were unable to bring these data into the RDC and instead the data file provided to us by NCHS through the RDC actually contained cause-of-death data already linked with the DEMS cohort and nested case-control data. Through the use of this combined data file at the RDC, we were able to identify IHD deaths and perform our analyses of REC and RPM exposure and IHD mortality. (IHD cases were defined based on the following ICD codes: ICD-7: 420, ICD-8 & 9:410-414.)

5.1.1 Original DEMS cohort data

As a first step in our research, in 2013 we requested and received the original DEMS lung cancer mortality cohort data from the National Cancer Institute (NCI) and the National Institute for Occupational Safety and Health (NIOSH).

DEMS is a cohort mortality study of 12,315 workers at 8 non-metal mining facilities exposed to diesel exhaust and was initiated by NIOSH and NCI in 1997 in response to health concerns.[6] The cohort study was originally intended to investigate the exposure-response relationship between diesel exhaust and lung cancer. The original DEMS file contains information on job titles and work dates as well as sex, date of birth, and partial information on race. It also identifies subjects with underlying causes of death attributable to both lung and esophageal cancers.

Importantly, the original DEMS file also provides time-varying, job-specific historical estimates of quantitative exposure to RPM and REC. Details of the exposure assessment in DEMS are described extensively elsewhere.[19-23] REC exposure values were estimated from REC exposure measurements from personal samples collected during the 1998–2001 DEMS surveys. These measurements were then used to estimate historical annual REC concentrations for each job by taking into account the effects of changes in mine-specific diesel exhaust-related determinants (e.g., diesel engine horsepower, emission controls,

and exhaust air rates from each mine). Because historical REC measurements were not available, the original NCI and NIOSH investigators estimated REC using historical CO measurements under the assumption that global changes in these determinants affected CO and REC similarly[21,24]. Annual average daily exposure was assigned to each year of active employment for each participant from the year of dieselization in each facility to end of follow-up.

5.1.2 Case-control study nested in the DEMS cohort included data on smoking

In order to adjust for cigarette smoking and other potential confounders in the DEMS cohort study, NCI researchers initiated a case-control study nested within the original DEMS cohort.[25] This nested case-control study included 198 lung cancer deaths and 562 control subjects, individually matched on mining facility, sex, race/ethnicity, and birth year (within 5 years). Smoking data were collected directly from next of kin interviews for all lung cancer cases and their matched deceased controls, or directly from living controls via telephone interviews. In 2014, we requested and received the nested case-control data from NCI to allow us to adjust our findings for potential confounding by cigarette smoking.

5.1.3 IHD mortality data

The only mortality information contained in the original DEMS data file is for deaths attributable to lung and esophageal cancers. In order to obtain information on IHD mortality, we requested additional data on cause of death from NCI.

5.2 IHD Mortality Analysis

5.2.1 Estimated Exposure-Response using Penalized Splines in Cox Models (Aim 1)

5.2.1.1 Methods

DEMS is a cohort mortality study of non-metal, underground and surface miners from eight facilities exposed to diesel exhaust.[6] The facilities were located in Missouri (one limestone), New Mexico (three potash), Ohio (one salt), and Wyoming (three trona). These facilities were selected because workers had low exposure to known occupational lung carcinogens (particularly silica, radon, and asbestos) but were exposed to extensive diesel engine usage. Historical RPM and surrogate diesel exhaust information was available to assist in development of quantitative estimates of past exposure.[19-22] All workers who were ever employed in a blue-collar job for at least 1 year after dieselization at the study facilities were eligible for study. Follow up time began at date of dieselization of each mine, ranging from 1947 to 1967, or the date of hire, whichever came later, and continued through 1997. The total sample size in the DEMS cohort was 12,315 workers. This analysis is restricted to white, male workers who were hired after the start of dieselization. The restriction by hire date was imposed for two reasons; first, to reduce selection bias from left truncation for the RPM analysis, and second, to take advantage of the exposure assessment for RPM which was only focused on the post-dieselization environment.

Demographic and work history information came from facility personnel records, including date of birth, sex, race/ethnicity, job titles and dates, prior employment, vital status, and next of kin. Information on race/ethnicity was unavailable for 64% of the workers. Race and ethnicity were coded to white/Hispanic or black. Missing race/ethnicity were classified as white/Hispanic, because, where race/ethnicity was known, 98% were white and/or Hispanic. Due to very few cases of IHD among black workers, we restricted this analysis to white/Hispanic or unknown race/ethnicity. In a sensitivity analysis to adjust for

cigarette smoking, we relied on smoking information (ever/never) from the 562 cohort members selected by incidence density-sampling as controls in a nested-case control study of lung cancer, some of whom later died from heart disease.[25] We imputed smoking 100 times using logistic regression in SAS proc MI and ran the main model adjusted for the imputed smoking variable.

Vital status was ascertained via matching with the National Death Index (NDI-plus) and the Social Security Administration death files. The 111 subjects who could not be matched to NDI- plus or Social Security Administration were considered alive until their last observed date and censored afterward. Cause of death was obtained from NDI-plus from 1979 to 1997 and before that from death certificates coded by a certified nosologist. Among the men hired after the start of follow up, 298 deaths due to IHD were identified, 84 (28%) of them occurring while at work or within 6 months after leaving work.

Individuals who had worked at more than one study facility were assigned to the facility where they had worked the longest. However, their exposure estimates included all jobs in each facility where they had worked.

Exposure assessment in DEMS has been described in detail elsewhere.[19-23] Briefly, diesel exhaust exposure resulted from ore extraction, haulage, maintenance, and personnel transport vehicles. To estimate diesel exhaust, REC exposure values were estimated from personal samples collected during the 1998–2001 DEMS surveys. These measurements were used to estimate historical annual respirable elemental carbon concentrations for each job by taking account of changes in determinants of mine-specific diesel exhaust over time, such as diesel engine horsepower, emission controls, and exhaust air rates at each mine. Because historical REC measurements were not available, these effects were estimated using historical CO measurements under the assumption that global changes in these determinants would have affected CO and REC similarly.

RPM, composed of mine and ore dust, was treated as a confounder in the lung cancer analyses of DEMS; the original exposure of interest was diesel exhaust. Quantitative levels of exposure to RPM were based on a total of 497 area and 504 personal samples,[19] including area samples taken by the DEMS IH team.[21] Underground jobs were assigned to one of three groups: 1) production and crushing; 2) haulage and travel ways; and 3) shop and office.

Facility, job group, and decade specific estimates of the annual average daily diesel exhaust and RPM exposures were assigned at the individual level based on work history and sampling data. Annual average daily exposure was calculated as the mean concentration of daily 8 hour time weighted averages in each calendar year. Time-varying estimates of cumulative exposure to RPM (mg/m^3-years) and elemental carbon ($\mu g/m^3$-years) were calculated for each worker. Average intensity of RPM and elemental carbon were calculated as the cumulative exposure divided by duration of employment.

Cox proportional hazards models were fit to estimate the effect of cumulative exposure as well as time-varying average intensity to diesel exhaust and RPM on IHD mortality. The cumulative and average intensity metrics for each type of exposure was defined as categorical variables. Models were stratified by state (Missouri, New Mexico, Ohio, and Wyoming) to allow for different baseline hazards. The following variables were included to control for confounding: location (surface/underground), the other exposure (RPM or REC), and birth-year. Models were stratified by location (surface/underground) to examine

effect modification. Confounding by smoking was examined in a sensitivity analysis using multiple imputation, as described above.

5.2.1.2 Results

Table 1: Hazard ratios (HR) for IHD by metrics of exposure to respirable elemental carbon (REC) and to respirable dust (RPM), including cumulative exposure and time weighted average (TWA) concentration. Data from DEMS cohort for workers first employed after the start of dieselization.

	0-9	10-19	20-39	40-159	160-639	640-1,279	1,280-2,559	≥2,560	P-trend [a]	EHR/unit [a]
					Cumulative exposure to REC (μg/m^3-year)					
Cum REC HR [b]	1.00	0.98	0.65	0.64	1.35	1.10	0.77	1.18	0.03	1.03×10^{-4}
95% CI		(0.56,1.72)	(0.35,1.20)	(0.29,1.41)	(0.61,3.00)	(0.51,2.35)	(0.36,1.65)	(0.56,2.48)		
Cases	45	29	29	39	54	29	38	37		
					Cumulative exposure to RPM (mg/m^3-year)					
Cum RPM	0-4	5-9	10-14	15-19	20-24	25-29	30-39	≥40		
HR [c]	1.00	1.00	0.96	1.09	1.42	1.06	1.13	1.25	0.14 [d]	5.85×10^{-3}
95% CI		(0.69,1.44)	(0.64,1.45)	(0.68,1.74)	(0.89,2.26)	(0.61,1.86)	(0.67,1.90)	(0.78,2.01)		
Cases	62	55	42	29	32	18	23	39		
					TWA exposure to REC (μg/m^3)					
Ave REC	0.0-0.9	1.0-1.9	2.0-3.9	4.0-15.9	16.0-31.9	32.0-63.9	64.0-127.9	≥128.0		
HR [c]	1.00	0.94	0.80	0.51	0.47	0.43	0.66	0.75	0.16	6.94×10^{-4}
95% CI		(0.58,1.53)	(0.48,1.34)	(0.29,0.90)	(0.26,0.85)	(0.23,0.83)	(0.35,1.24)	(0.39,1.44)		
Cases	45	40	24	17	22	40	36	76		
					TWA exposure to RPM (mg/m^3)					
Ave RPM	0.0-0.49	0.50-0.69	0.70-0.99	1.00-1.49	1.50-1.99	2.00-2.49	2.50-2.99	≥3.00		
HR [b]	1.00	1.08	1.57	1.52	2.27	2.11	2.24	2.58	0.05 [d]	0.413
95% CI		(0.59,1.96)	(0.93,2.64)	(0.80,2.87)	(1.21,4.28)	(1.03,4.29)	(1.03,4.86)	(1.26,5.28)		
Cases	29	31	52	38	47	36	20	47		

[a] Change in HR per unit increase in exposure

[b] For REC metrics, score test/estimated slope parameter of linear trend beyond change-point in segmented regression: 230 μg/m^3-years for cumulative REC and 2.0 μg/m^3 for TWA REC (see text). Model stratified on mine type and adjusted for birth year and cumulative REC and TWA RPM.

[c] Model stratified on mine type and adjusted for birth year and TWA REC and cumulative RPM.

[d] Score test of no trend.

The primary results from the standard analysis are presented in Table 1. The categories were defined to divide the exposure range into equal segments on the log scale. As seen above, the HR and 95% CI for the top quantile versus lowest category of exposure were 1.18 (95% CI: 0.56, 2.24) for cumulative REC, 1.25 (95% CI: 0.78, 2.01) for cumulative RPM, 0.75 (95% CI: 0.39, 1.44) for average intensity REC and 2.58 (95% CI: 1.26, 5.28) for average RPM. The test for trend across the increasing exposure categories was significant for cumulative exposure to REC, although the absolute HRs were below the null for the first several categories until 160 $\mu g/m^3$-years. The trend was also significant for average annual exposure intensity for RPM, beginning at 1.08 and rising to 2.58 in the highest category. These results provide stronger evidence for the association between risk of IHD mortality and RPM than exposure to REC.

5.2.2 Examine the effects of different windows of exposure (Aim 2)

5.2.2.1 Methods

Sensitivity analyses were also conducted to assess whether the effect of exposure was stronger among people still at work and currently experiencing exposure. In this analysis, the person-time was restricted to active working time with a 6-month buffer (i.e., employed or employment terminated less than 6 months ago). In order to assess the impact of recent exposure on risk of IHD, we conducted analyses of cumulative diesel and respirable particulate matter exposure restricted to IHD deaths that occurred while the worker was still at work or within 6 months of leaving work.

5.2.2.2 Results

We saw a slightly stronger increase in risk from cumulative REC exposure among those actively employed compared to the full cohort (Figure 1). The risk for IHD from cumulative RPM was much stronger among those who were actively employed, HR = 1.015 (95% CI: 1.003, 1.027) compared to all workers (Figure 2).

Figure 1: HR for IHD mortality and cumulative REC for actively employed (dashed) versus the entire cohort (solid)

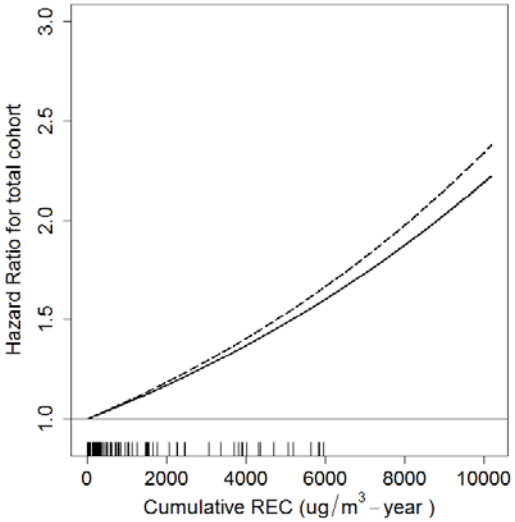
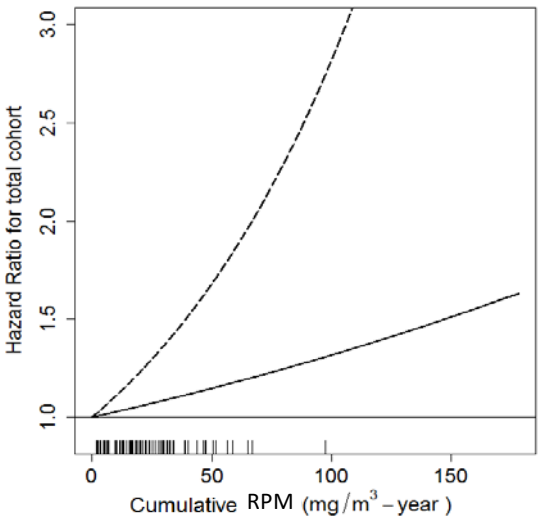


Figure 2: HR for IHD mortality and cumulative RPM for actively employed (dashed) versus the entire cohort (solid)



5.2.3 Parametric g-formula to address time-varying confounding i.e, HWSE (Aim 3&4)

5.2.3.1 Methods

Although we proposed to apply g-estimation, we chose an alternative g-method, parametric g-formula instead because it is easier to incorporate quantitative exposure. For the current analysis, workers employed in a blue collar job at a selected mine for at least 1 year after dieselization were eligible. We limited the study population to workers who ever worked underground (n=8,307) and therefore omitted all person-time (and IHD events) that occurred prior to a worker's first underground job. We identified 302 IHD deaths among the ever-underground subcohort. Since all cases except one were male, analyses in the current study were restricted to the 8,063 male workers.

Table 2: Characteristics of a sub-cohort of US male non-metal miners who ever worked underground

Characteristic	No	%	Mean (SD)
No. of participants	8,063		
White and/or Hispanic	7,935	98.4	
Total deaths	1,390		
IHD deaths	301		
Mine Type/State			
Limestone/Missouri	938	11.6	
Potash/New Mexico	3,186	39.5	
Salt/Ohio	474	5.9	
Trona/Wyoming	3,465	43.0	
Prevalent hires	556	6.9	
Age at baseline			30.5 (9.2)
Age of IHD cases			65.0 (11.4)
Duration of follow-up			21.6 (10.1)
Duration of active employment			11.9 (9.5)

During our lung cancer analysis, we fit an accelerated failure time model and were able to demonstrate that termination of employment is associated with exposure in this cohort (See Table 6). This motivated our use of g-methods for this analysis, which are equipped to deal with time varying confounding affected by previous exposure. We applied the parametric g-formula to assess the risk of IHD mortality under a simulated natural course (i.e., what actually happened) and various hypothetical interventions limiting exposure to REC (including no exposure and limits based on ACGIH current and past recommendations), as well as joint interventions on REC and RPM. We fitted parametric models for the outcome (IHD mortality), competing events (non-IHD mortality) and all time-varying covariates (employment status, job location) and exposures (REC and RPM), conditional on prior exposure and covariate histories as well as baseline covariates (age, calendar year, race/ethnicity, state and an indicator variable for workers hired prior to dieselization).

We then generated a pseudo-sample based on the observed distributions of the baseline covariates of equal size to observed data (n=8,063). In this pseudo-sample we simulated exposure and covariate values at each age using the parameters of the models for the exposure and covariates until age 90 years. Under no intervention, the simulation used values for the exposure that are predicted from the model and then predicts the risk under the natural course of events. For all other interventions, the exposure values were changed from the predicted values according to the specific intervention. The values for all covariates at the subsequent time point, including probabilities of death due to competing risks and due to IHD, were

then predicted using the simulated exposure and covariate values at each age and the parameters from the covariate, outcome, and competing risks models.

Cumulative incidence of IHD, with age as the time scale, was calculated for each intervention using an estimator for the sub-distribution of the event of interest, in the presence of competing risks.[26] We estimated 95% CIs by repeating the entire above process in 200 bootstrap samples. The SD of the estimates from the bootstrap samples was used as an estimate of the standard error to generate the 95% CI. Sensitivity analyses drawing larger Monte-Carlo pseudo-samples (N=50,000) to assess stability of results were conducted.

We also repeated analyses using our imputed smoking data (see 5.4 below). Smoking data were available for 361 participants in our primary analysis subset (4.5%) and analyses with imputations for smoking status (ever-never) only as well as smoking status *and* intensity (4-level categorical variable) were performed with the imputed smoking variables entered as baseline covariates in the prediction models. We calculated and averaged estimates of cumulative incidence of IHD mortality for selected interventions in 100 imputed datasets generated using the MI procedure in SAS. All analyses were carried out in SAS (SAS version 9.4; SAS Institute Inc., Cary, NC).

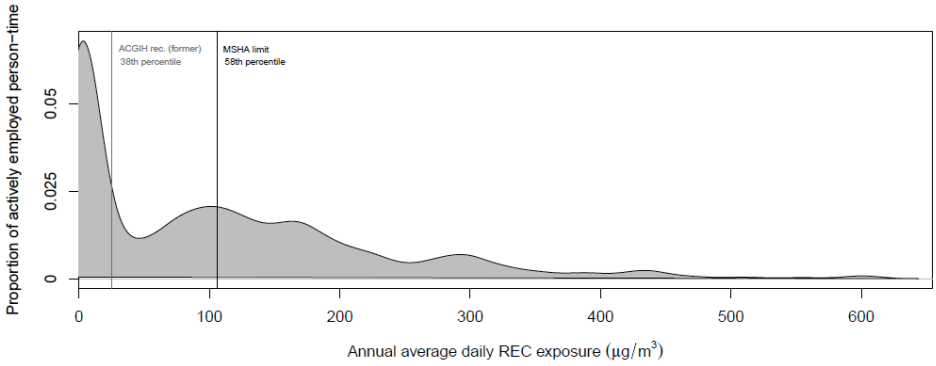
Counterfactual risk was estimated under the natural course (i.e., no intervention) and a series of hypothetical interventions. In total, we estimated risk under four exposure scenarios: no intervention, always unexposed (occupational REC set to 0 $\mu g/m^3$ for everyone), and two different maximum annual average daily REC exposures. The MSHA exposure limit for diesel exhaust is 160 $\mu g/m^3$ total carbon (TC) measured on the submicrometer particulate matter fraction.[27] We used internal exposure assessment data from DEMS [20-22] to convert TC limits to appropriate EC values and subsequently EC values from the submicron level to the appropriate respirable fraction. The equivalent value for 160 $\mu g/m^3$ of TC was 106 $\mu g/m^3$ of REC in DEMS. The two exposure limits evaluated were based on the current MSHA limits of 106 $\mu g/m^3$ REC, and the American Conference of Governmental Industrial Hygienists (ACGIH) recommendation of 25 $\mu g/m^3$ REC (later retracted).[28]

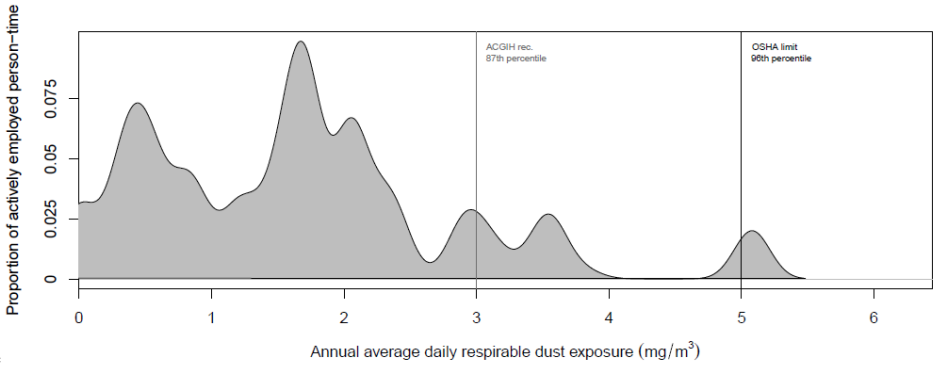
We also estimated counterfactual risk under joint interventions on both REC and RPM, using the current Occupational Safety and Health Administration (OSHA) regulation and the ACGIH recommendation. The three joint interventions were for maximum limits of a) 106 $\mu g/m^3$ REC and 5 mg/m^3 RPM (OSHA regulation); b) 25 $\mu g/m^3$ REC and 3 mg/m^3 (ACGIH recommendation); and c) 0 $\mu g/m^3$ REC and 0.5 mg/m^3 since an intervention of no-RPM exposure was not considered feasible in this occupational setting.

5.2.3.2 Results

The exposure distribution for REC and RPM among actively employed person-time is depicted in Figure 3, with indicators for the percentile of the exposure distribution corresponding to the exposure limits in each hypothetical intervention. The REC equivalent for the current MSHA regulatory limit in this study corresponded to the 58th percentile of the observed REC exposure distribution, while the equivalent for the ACGIH recommended value corresponded to the 38th percentile. By contrast, the current OSHA regulatory limit for RPM corresponded to the 98th percentile of observed RPM exposures, with the ACGIH recommended value corresponding to the 87th percentile. Exposure at 0.5 mg/m^3 corresponded to the 20th percentile of RPM exposure.

Figure 3: Annual average daily REC (in $\mu g/m^3$) and respirable dust (RPM in mg/m^3) exposure distributions during active employment. Vertical lines represent values of maximum limits set in hypothetical interventions assessed in this study.





*

Interventions on REC exposures resulted in reduced risk of IHD mortality with lower (stricter) limits corresponding to greater reductions in risk. Joint interventions for REC and RPM resulted in slightly bigger reductions in risk with the risk ratio (RR) for the most stringent REC and RPM joint intervention reaching statistical significance (RR = 0.81, 95% CI: 0.69, 0.96). The RR comparing the risk under the hypothetical intervention setting a maximum REC exposure to 106 µg/m^3 and the natural course was 0.92 (95% CI: 0.84, 1.00), while the RR comparing always unexposed to REC and the natural course was 0.85 (95% CI: 0.65, 1.11). The risk difference (RD) for the intervention comparing always unexposed to REC and the natural course was -2.9 (95% CI: -7.6, 1.7). Based on these estimates the excess fraction (AF = (RR - 1)/RR)[31] of IHD deaths due to the observed diesel exhaust exposures in this population was estimated at 15%. The risk difference comparing a joint intervention for always unexposed to REC and setting a RPM exposure maximum of 0.5 mg/m^3 was -3.6 (95% CI: -6.3, -0.9). Cumulative incidence graphs under the natural course and the hypothetical interventions are depicted in Figure 4, with increasing reduction in risk over time (age in years) when comparing interventions to the natural course.

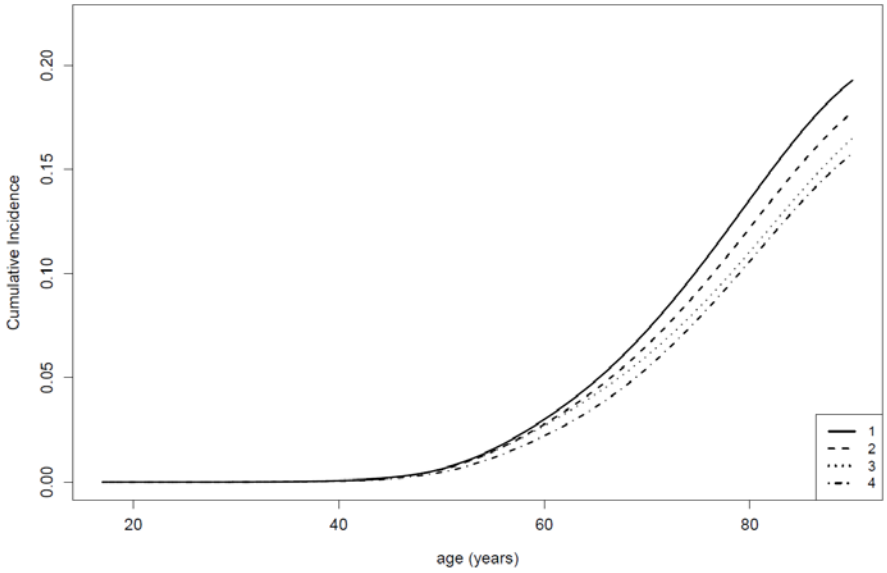
Table 3: Risk of IHD mortality under the natural course and under hypothetical interventions on REC and RPM

Intervention	IHD mortality %	RR	95% CI	Risk Diff.	95% CI
Natural course (no intervention)	19.4	.	.	.	.
Interventions on REC only					
REC≤106 µg/m^3	17.9	0.92	0.84, 1.00	-1.5	-3.0, 0.0

REC≤25 μg/m^3	16.9	0.87	0.72, 1.06	-2.5	-6.0, 1.0
REC=0 μg/m^3	16.5	0.85	0.65, 1.11	-2.9	-7.6, 1.7
Interventions on RPM					
RPM ≤5 mg/m^3	19.4	1.00	1.00, 1.00	0.0	0.0, 0.0
RPM ≤3 mg/m^3	19.3	0.99	0.98, 1.00	-0.1	-0.4, 0.2
RPM ≤0.5 mg/m^3	18.8	0.97	0.84, 1.11	-0.6	-3.9, 2.0
Joint interventions on REC & RPM					
REC≤106 μg/m^3 & RPM ≤5 mg/m^3	17.9	0.92	0.84, 1.02	-1.5	(-2.9, 0.2)
REC≤25 μg/m^3 & RPM ≤3 mg/m^3	16.8	0.87	0.73, 1.03	-2.6	(-5.7, 0.5)
REC=0 μg/m^3 & RPM ≤0.5 mg/m^3	15.8	0.81	0.69, 0.96	-3.6	(-6.3, -0.9)

IHD, ischemic heart disease; REC, respirable elemental carbon; RPM, respirable particulate matter

Figure 4: Cumulative incidence of IHD mortality under the natural course (solid line) and under three hypothetical interventions: 1) simulated natural course, 2) REC<=106 μg/m^3, 3) REC=0 μg/m^3 and 4) REC=0 μg/m^3 & RPM <=0.5 mg/m^3.



Incorporating smoking data into the analysis of IHD mortality (Aim 4)

Results from sensitivity analyses with larger Monte-Carlo samples did not differ significantly from the primary results. Results from multiply imputed datasets for missing smoking information also resulted in similar estimates to the primary analysis results for the same interventions for REC exposures (Table 4).

Table 4: Risk ratios comparing IHD mortality under the natural course and under hypothetical interventions on REC, adjusting for smoking using imputed data on smoking status and smoking status and intensity.

Intervention	Adjusting for smoking status		Adjusting for smoking status and intensity	
	Avg. Risk Ratio*	Range**	Avg. Risk Ratio*	Range**
REC≤106 μg/m^3	0.93	0.86 – 1.01	0.94	0.90 – 0.97
REC≤25 μg/m^3	0.88	0.79 – 0.94	0.88	0.81 – 0.92
REC=0 μg/m^3	0.86	0.73 – 0.94	0.86	0.78 – 0.91

*Average risk ratios from 100 multiply imputed datasets

**The range in risk ratios from the 100 multiply imputed datasets.

5.2.3.3 Conclusions

In this analysis, we were able to provide an unbiased estimate of risk of IHD mortality in relation to diesel exhaust, specifically REC, by controlling for the healthy worker survivor effect, by applying the parametric g-formula in the DEMS subcohort of 8,063 ever-underground miners. Based on these results we evaluated the effects of plausible (hypothetical) interventions to reduce exposures. We estimated reduced cumulative IHD mortality risk with hypothetical interventions that set historical exposures in this cohort to limits according to current regulations/guidelines, and found that greater reductions in IHD mortality were associated with lower exposure limits.

Joint interventions on REC and RPM resulted in only a slightly greater reduction in risk than interventions on REC alone. It should be noted, however, that the much smaller contribution of RPM exposures to IHD risk may be due in part to greater exposure misclassification that may exist for this exposure than REC exposure in DEMS. Estimates of personal RPM exposure in DEMS were based on less than 300 measurements in eight mines over 30 years.[19] Exposure assessment for REC by comparison, was based on thousands of measurements and a much more detailed protocol for modelled historical exposure levels,[19, 20, 21] and although still subject to exposure misclassification, is considered more accurate.

Risk estimates comparing IHD mortality under hypothetical interventions on REC exposure to IHD mortality under the natural course were smaller in magnitude on the relative scale than our estimates of the impact from the same hypothetical exposure interventions on lung cancer (see section 5.3 below). Given the higher prevalence of IHD mortality, however, risk reduction associated with exposure interventions was greater on the absolute scale, i.e., the number of deaths, which is a more meaningful measure of public health impact. The number of IHD deaths prevented due to a beneficial intervention on exposure would be greater than the number of lung cancer deaths prevented by the same intervention despite the apparent weaker exposure-response relationship on the relative scale.

5.3 Lung Cancer Re-Analysis, adjusting for HWSE (Aim 5)

In our re-analysis of the original DEMS cohort lung cancer data, we employed the same g-methods described in 4.2.3 to adjust for the HWSE.

5.3.1 Methods

The population for this analysis was restricted to miners in the cohort who ever worked underground (n=8,307), within which 122 lung-cancer deaths were ascertained. Table 5 below summarizes demographic characteristics in the DEMS ever-underground subcohort. The cohort is predominantly male and, among those with information on race/ethnicity, 98% of participants were white and/or Hispanic. The median year at start of follow up was 1975, with mean (±SD) duration of active work history of 20 (±9.5) years. The mean age (± SD) at baseline was 30 (±9.1) years, while age at death due to lung cancer ranged from 44 to 90 years of age.

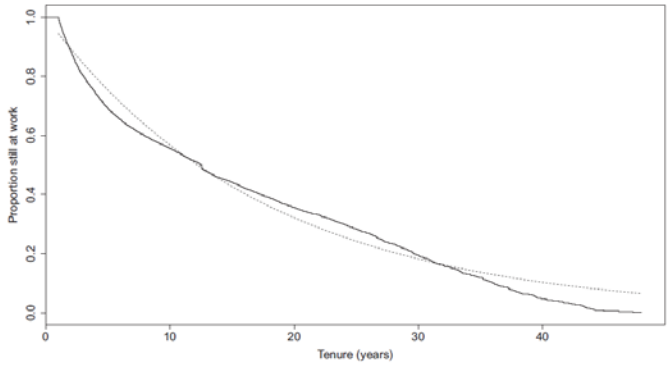
The mean (SD) annual average daily REC exposures in the ever-underground subcohort was 111.5 (117.7) $\mu g/m^3$, with a range of 0 to 604.3 $\mu g/m^3$ and interquartile range width (IQR_w) of 167.3 $\mu g/m^3$.

Table 5: Demographic Characteristics at Baseline in the DEMS ever-underground sub-cohort

No. of participants	8,307
Male, no. (%)	7,478 (97)
Race, no. (%)	
Unknown	3,773 (49)
White/Hispanic	3,776 (49)
Black	127 (2)
Mine type, no. (%)	
Limestone	946 (11)
Potash	3,278 (40)
Salt	497 (6)
Trona	3,586 (43)
Lung cancer deaths, no.	122
Total deaths, no.	1,404

Because no diesel exhaust was present in mines prior to the start of follow-up, left-truncation bias did not need to be addressed in REC analyses, but when examining the joint effects of REC and RPM, adjustments for left-truncation bias were made by controlling for prior work experience. To evaluate whether employment termination was affected by the exposure, in which case traditional regression would give biased results, we fitted an accelerated failure time model to estimate the effect of REC exposure on time to termination of active employment. We assumed an exponential distribution of time to termination as shown in Figure 5.

Figure 5: Time to termination of employment curve in the DEMS ever-underground subcohort (*solid line*), compared with a parametric exponential survival function with hazard $\lambda = 0.06$ (*dashed line*).



For this model, the data were truncated to only include person time with active work history. We defined termination of active employment date as the last observed date of active work before December 31, 1997, for those alive as of December 31, 1997. Those with December 31, 1997 as their last observed active work date were considered administratively censored at that time. Death during active work was the only censoring event. The model included annual average daily exposure to REC as continuous variables, as well as sex, race/ethnicity, facility, birth year, and a natural cubic spline for calendar year. The process was repeated restricting to incident hires, where follow-up time was essentially equal to tenure as a miner post-dieselization. As sensitivity analyses, subsets of the population with at least 3 and 5 years tenure after dieselization were examined. Given the exponential distribution of termination times as

indicated in Figure 5, the inverse of the time to termination ratios obtained by the accelerated failure time model can be interpreted HRs as for the risk of termination of active work. The relationship between time to termination of employment and lung cancer mortality was assessed using a Cox regression model, controlling for sex, race/ethnicity, state, and cumulative exposure, with attained age as the time scale.

Once we had determined that previous exposure affected the time-varying confounder employment status, we applied the parametric g-formula to estimate the effect of exposure interventions on the risk of lung cancer. We assessed the effect of hypothetical limits on diesel exhaust in terms of cumulative risk of lung cancer mortality. In total, we estimated risk under six exposure scenarios: no intervention, always unexposed (occupational REC set to 0 $\mu g/m^3$ for everyone), and four different maximum annual average daily REC exposures. The four exposure limits evaluated were based on the past and current MSHA-equivalent limits of 260 and 106 $\mu g/m^3$of REC,[27] and the past and current ACGIH-equivalent recommendations of 65 and 25 $\mu g/m^3$ REC.[28]

Under the assumptions of (1) no unmeasured confounders (i.e., conditional exchangeability) at all time points, (2) counterfactual consistency (i.e., every individual's counterfactual outcome under their observed exposure history is equal to their observed outcome), and (3) correct model specification, the parametric g-formula is a generalization of standardization for time-varying exposures and covariates.[17, 29] The risk of lung cancer under each intervention is estimated as a weighted sum (or integral), over exposure and covariate histories, of the probability of lung cancer conditional on exposure and covariates.

Our analysis estimating lung cancer risk under hypothetical REC limits using the parametric g-formula involves parametric models for the outcome and all time-varying covariates and exposures, conditional on prior exposure and covariate histories as well as baseline covariates. A Monte Carlo estimator approximates the integral or weighted sum of covariate and exposure histories. This is achieved by generating a pseudo-sample based on the observed sample population. Parametric models were fitted for the outcome (lung cancer death), competing risk (non-lung cancer deaths), exposure (annual average daily REC levels), and time-varying covariates for active employment status and job location (with different levels for underground, surface, and inactive), all conditional on prior covariate and exposure histories and baseline covariates (age, calendar year, sex, race/ethnicity, state, and indicator for workers hired before dieselization). Person-time used in these models was restricted to that accrued before and including 90 years. Both REC exposure and employment status were lagged by 15 years in outcome models to account for cancer latency. Current annual average daily REC exposure values as well as cumulative exposure up to the previous year were also included in outcome and competing risk models.

In the subsequent pseudo-sample of the same size as the observed ever-underground subcohort (n = 8,307), each worker was followed from their age at start of follow-up until age 90 years. We performed the same process with a larger pseudo-sample (n = 50,000) as a sensitivity analysis. Exposure and covariate values at each age (years) were simulated using the parameters of the models for the exposure and covariates from above. Under no intervention, the simulation uses values for the exposure that are predicted from the model and then predicts the risk under the natural course of events. For the other interventions, the exposure values were changed from the predicted values according to the value of the specific intervention. The covariate values for the subsequent time point, including probabilities of death due to competing risks and due to lung cancer were then predicted using the simulated exposure and covariate values at each age and the parameters from the covariate, outcome, and competing risks models.

We then calculated cumulative incidence of lung cancer, with age as the time scale, for each intervention using a cumulative incidence estimator for the sub-distribution of the event of interest, in the presence of competing risks.[26] We obtained CIs for cumulative incidences and RRs by repeating the above process in 200 bootstrap samples. Statistical analysis involving accelerated failure time models for termination of

active work was performed using R software (version 3.0.2), while the parametric g-formula was performed in SAS (SAS version 9.4; SAS Institute Inc., Cary, NC).

5.3.2 Results

HR and 95% CI for an IQR_w increase in REC exposure and termination of employment are summarized in Table 6. An IQR_w increase in REC exposure was associated with 36% (95% CI = 33%, 39%) shorter time to termination of active work. Results were similar after restriction to incident hires. The effect was attenuated after analysis was restricted to incident hires with at least 3- and 5-year tenures.

Table 6. HR with 95% CI from conditional accelerated failure time models for an interquartile range Width (IQR_w = 167 ug/m^3) increase in REC exposure and termination of active employment, in the DEMS ever-underground subcohort

Subset	HR (95% CI)[a]
All subjects (n = 8,307)	1.6 (1.5, 1.6)
Incident hires[b] only (n = 7,750)	1.5 (1.4, 1.6)
Incident hires with ≥3 years tenure (n = 5,993)	1.3 (1.3, 1.4)
Incident hires with ≥5 years tenure (n = 4,971)	1.2 (1.2, 1.3)

[a]This approximates the HR for the event associated with an IQRw increase in exposure at time *t*. Assuming exponential distribution of time to termination then the HR is the inverse of survival time ratio obtained by the accelerated failure time model.
[b]Incident hires are defined as those miners who started working at the participating mines after dieselization.

The majority of deaths (80%) and lung cancer deaths (91%) occurred after termination of active work history. Workers with longer employment had lower risks of lung cancer, with a HR of 0.97 (95% CI = 0.94, 0.99) for each additional year of active work history (lagged by 15 years) estimated in Cox models adjusted for baseline covariates and cumulative exposure, using attained age as the time scale of interest.

The observed cumulative incidence of lung cancer among ever-underground workers is compared with the simulated cumulative incidence under no intervention in Figure 6. Figure 7 shows the simulated cumulative incidence under all proposed interventions. RRs comparing the risk under each intervention to the simulated risk under no intervention are presented in Table 7. The lifetime cumulative incidence when everyone was always unexposed to REC was estimated at 5.7%, for a risk difference of 2.1% compared with the observed risk of 7.8%. The attributable fraction (AF = (RR - 1)/RR)[30] of lung cancer deaths due to the observed diesel exhaust exposures in this population is estimated at 27%. Results did not substantially differ with a larger pseudo-sample simulation.

Figure 6: Observed (*solid line*) and simulated (*dashed line*) cumulative lung cancer mortality in the DEMS ever-underground cohort, under the natural course.

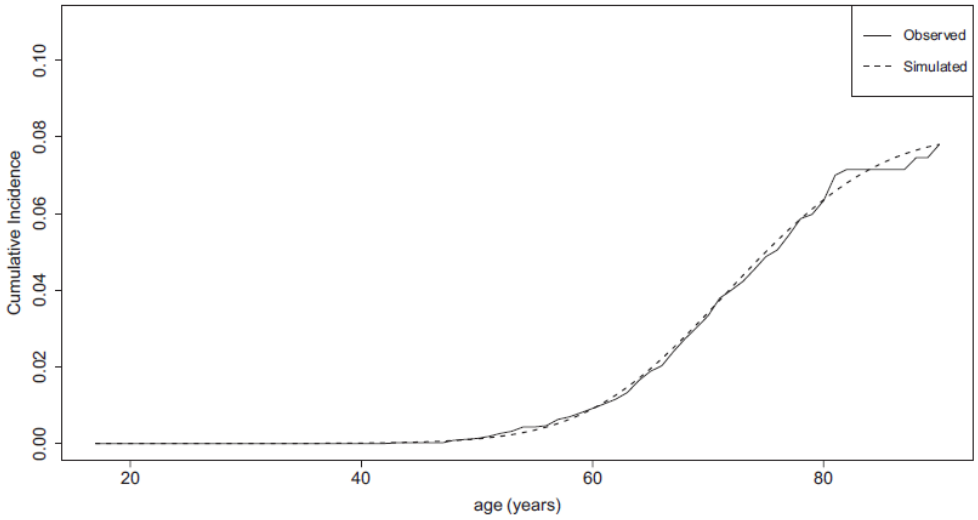


Figure 7: Simulated cumulative lung cancer mortality in the DEMS ever-underground cohort, under no intervention (*solid line*) and multiple interventions (*dashed lines*) regarding limits of maximum average annual daily REC.

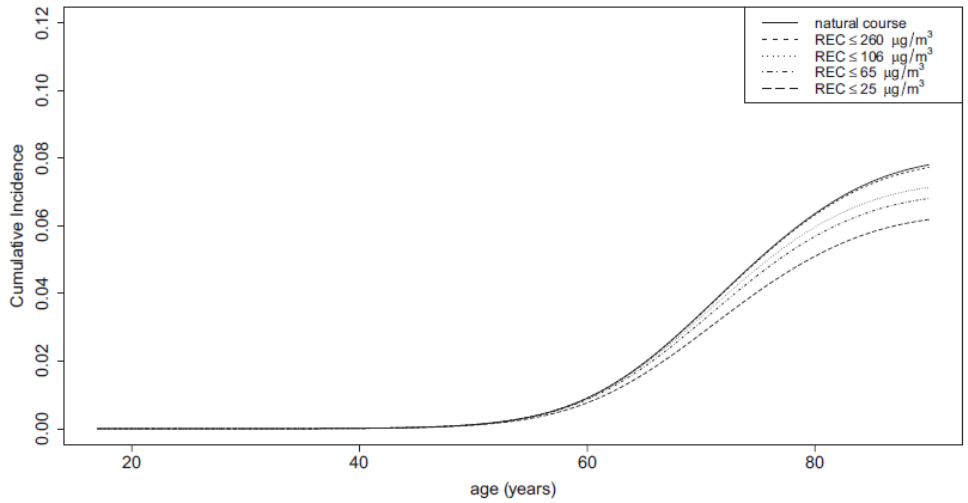


Table 7: Cumulative incidence for lung cancer mortality and corresponding RR and 95% CI, under several interventions in the DEMS ever-underground cohort

Intervention	Cumulative Incidence	RR (95% CI)	% Person-time Above Intervention Limit	% Active Work Person-time Above Intervention Limit
Observed data	7.8			
Simulated natural course[a]	7.8			
Max REC limit at ≤260 μg/m^3 [b]	7.7	1.0 (0.9, 1.0)	5	10
Max REC limit at ≤106 μg/m^3 [b]	7.1	0.9 (0.7, 1.2)	20	39
Max REC limit at ≤65 μg/m^3 [b]	6.8	0.9 (0.6, 1.2)	28	53
Max REC limit at ≤25 μg/m^3 [b]	6.2	0.8 (0. 5, 1.1)	33	63
Always unexposed[c]	5.7	0.7 (0.4, 1.1)	50	96

The interventions were as follows:
[a]the simulated natural course simply simulates what actually happened under no intervention, where everyone receives exposure and covariate histories predicted given the observed data.
[b]Simulations with interventions with max REC limits, replaced REC values predicted above the limit, with the specified intervention limit.
[c]Under always unexposed REC is set to 0 at all times.
REC indicates respirable elemental carbon.

Using the g-formula, we also computed the exposure level necessary to limit excess risk in this population to the common regulatory action level of one lung cancer death per 1,000 workers. Similar to conventional risk management analyses, we assumed that all workers were continuously exposed between ages 20 and 65 years. Using the observed background rates in the study population, we predict that REC exposure would need to be limited to below 1 μg/m^3 to achieve this level of excess risk.

5.3.3 Conclusions

In this analysis, we examined a factor that could lead to an underestimate of risk of lung cancer – the healthy worker survivor bias – and applied the parametric g-formula in the DEMS subcohort of 8,307 ever-underground miners. We estimated lifetime risk of lung cancer under different hypothetical interventions targeting REC exposure, and saw continued diminishing risk of lung cancer with REC levels below the current MSHA exposure limits.

We evaluated the potential for healthy worker survivor bias using an accelerated failure time model, and observed an effect of exposure on time to termination of employment. Harmful exposures leading to termination of employment will typically result in a downward bias in the effect of cumulative exposure on the outcome because workers most susceptible to the outcome may leave work and accumulate less exposure. In this analysis, the effect of REC exposure was attenuated among those with at least 5 years of tenure, consistent with previous results reported by Attfield et al.[6], showing strong dose–response after among those with at least 5-year tenure. Bias in estimates of relative risk of lung cancer from diesel exposure due to exposure and termination of employment associations is likely less of an issue in this subset of the population.

We implemented the parametric g-formula to assess the effect of hypothetical interventions to reduce REC levels. These hypothetical interventions specified in our analysis could plausibly be implemented in real-world settings, and the parametric g-formula allows us to estimate effects of such interventions using observed data. In addition, under the assumptions of conditional exchangeability, consistency, and correct model specification also required in other g-methods and implicit in all analyses of observational data, the parametric g-formula allows us to adequately control for time-varying confounding affected by past exposure.

Interventions based on the intervention limits should lower REC concentrations so that they do not exceed the proposed limits in the working areas of all potentially exposed miners. This could be accomplished through any combination of emissions controls (such as increased ventilation) or actual reduction in emissions (through the use of cleaner fuel or newer engines). We expect that any such combination of

methods would have the same effect, an assumption required in order for counterfactual consistency to hold.

In this study, we found that an exposure of less than 1 $\mu g/m^3$ REC would be needed to achieve a lung cancer risk of 1 in 1,000. Although this figure is only an estimate, it suggests that a health-based exposure limit would make it difficult to use diesel equipment underground.

Under hypothetical interventions which set limits ranging from the past and current MSHA limits to ACGIH recommendations, we found that estimated lifetime risks of lung cancer were reduced with increasingly lower limits. Using the parametric g-formula in a reanalysis of the DEMS ever-underground subcohort, we observed results consistent with previous studies of diesel exhaust exposure and lung cancer risk using quantitative measures of exposure to elemental carbon,[25, 31] which found excess lung cancer risk in the lower observed ranges (1–25 $\mu g/m^3$) of occupational elemental carbon exposures.

5.4 Adjustment for Smoking

5.4.1 *Methods*

Smoking status was determined for each subject in the nested case-control study (n=760) through interviews with the subject (if alive) or next-of-kin (if deceased), but this represents only a small percentage of the full DEMS cohort population (n=12,315). We used smoking information from the nested case-control study subjects to adjust for confounding by smoking in the full cohort using multiple imputation. This entails the creation of a large number of full data sets (M=100) in which we imputed values of the smoking variable for the subjects in the full DEMS cohort population whose smoking status was unknown. Then we performed the analysis in parallel in each of the M data sets and combined the results together to get our final result. The variance estimates for this final result accounts for the uncertainty in each subject's true smoking status.

We used two different approaches for our multiple imputation. For both approaches, we performed the imputation with the 'PROC MI' procedure in SAS and combined the estimates using either the 'proc mianalyze' procedure or Rubin's rules1, as appropriate.

> Method 1: We have created a person-year database containing all fully observed data (with smoking) and partially observed data (without smoking). Where possible, each observation has been coded as either current, past, or never smoker. We imputed the values of the missing smoking using this same categorical variable. Although this results in inconsistent smoking histories in the imputed data (a current smoker one year may be a never smoker the next), it reflects the uncertainty about the smoking status of this worker.
>
> Method 2: We created a person-level database where smoking is coded with two variables: (1) an indicator for smoking at baseline and (2) a continuous measure of time-until-cessation of smoking. This imputation is a two-stage process. Smoking status is imputed first, and then time-until-cessation is imputed among those workers coded as smokers. We used this information to create coherent longitudinal histories of smoking for those analyses that require time-varying measures of the confounder.

5.4.2 *Results*

Although we used multiple imputation methods to adjust for potential confounding in our analysis of IHD mortality using the parametric g-formula, the scarcity of smoking data from the nested case-control study

limited our ability to properly account for smoking as a potential confounder. Results from analyses using imputed smoking information based on smoking data from the nested case-control study indicate that smoking may not be a strong confounder for IHD mortality in this setting, but these results are based on the very limited portion of the cohort with existing smoking data.

6.0 Dissemination Efforts and Highlights

We have focused on disseminating the analytical approaches we used to provide non-biased estimates of the exposure-response relationships between REC, RPM and both lung cancer and IHD mortality. To this end, we have produced manuscripts and given presentations at international conferences to share our findings with a broad range of stakeholders, including researchers, labor and inRPMry representatives. Below is a list of our manuscripts and conference presentations.

6.1 Peer reviewed publications:

Neophytou AN, Picciotto S, Costello S, Eisen EA. Occupational Diesel Exposure, Duration of Employment, and Lung Cancer: An Application of the Parametric G-Formula. *Epidemiology.* 2016 27(1):21-28.

Neophytou A, Picciotto S, Costello S, Eisen EA. Rejoinder Re: "Exposure to diesel motor exhaust and lung cancer risk: Are the calculated exposure limits really reliable?"
Epidemiology 2017 (In Press).

6.2 Manuscripts in preparation:

Neophytou AN, Costello S, Picciotto S, Brown DM, Attfield MD, Blair A, Steward PA, Lubin JH, Vermeulen R, Silverman DT, Eisen EA. "Occupational diesel exhaust and RPM exposures and risk of ischemic heart disease: an application of the parametric g-formula in a mining cohort." To be submitted to EHP for publication, May 2017.

Costello S, Attfield MA, Neophytou AM, Picciotto S, Brown DM, Attfield MA, Blair A, Stewart PA, Lubin JA, Vermeulen R, Silverman DT, and Eisen EA. "Ischemic heart disease from diesel exhaust and RPM exposure among underground, non-metal miners in the United States" To be submitted to EHP for publication, May 2017.

6.3 Posters and presentations:

Neophytou AN, Picciotto S, Costello S, Eisen EA. "'The Diesel Exhaust in Miners Study: Applying the Parametric G-formula to Evaluate Interventions for Occupational Risk Assessment." Invited Oral Presentation. Joint meeting of the West Virginia Coal Mining Institute, West Virginia Coal Association and the Central Appalachian Section of the Society for Mining, Metallurgy & Exploration (SME), Roanoke, West Virginia, May 2015.

Neophytou AN, Picciotto S, Costello S, Eisen EA. "Applying a structural nested accelerated failure time model for the effect of exposure on time to termination of employment: assessing healthy worker survivor bias in the Diesel Exhaust in Miners Study." Poster Presentation. Society of Epidemiology Research (SER) Annual Meeting, Denver, CO, June 2015.

Neophytou AN, Picciotto S, Costello S, Eisen EA. "Applying a structural nested accelerated failure time model for the effect of exposure on time to termination of employment: assessing healthy worker survivor

bias in the Diesel Exhaust in Miners Study." Oral Presentation. International Society of Environmental Epidemiology (ISEE) Conference, Sao Paulo, Brazil, August 2015.

Costello S, Neophytou AN, Brown DM, Attfield MD, Blair A, Steward PA, Lubin JH, Vermeulen R, Silverman DT, Eisen EA. "Ischemic heart disease from diesel exhaust exposure among underground, non-metal miners in the United States." Oral Presentataion. Epidemiology meeting of the International Congress on Occupational Health (EPICOH) Conference, Barcelona, Spain, September 2016.

Neophytou AN, Costello S, Picciotto S, Brown DM, Attfield MD, Blair A, Steward PA, Lubin JH, Vermeulen R, Silverman DT, Eisen EA. "Estimating risk of ischemic heart disease mortality under hypothetical interventions on occupational diesel exposures in a cohort of underground miners." Poster Presentation. International Society of Environmental Epidemiology (ISEE) Conference, Rome, Italy, September 2016.

Costello S, Attfield M, Lubin J, Neophtou AN, Blair A, Brown D, Stewart P, Vermeulen R, Eisen EA, Silverman D. Ischemic heart disease mortality, diesel exhaust and particulate matter exposure in the Diesel Exhaust in Miners Study (DEMS). Submitted for oral presentation at EPICOH Conference, Edinburgh, Scotland. September 2017.

7.0 Conclusions and Impact Assessment

We found that *exposure to REC* was associated with *shorter time to termination of employment*, demonstrating that results will misleadingly suggest that more exposure reduces risk of IHD if healthy worker survivor effect is not addressed using appropriate statistical methods. The statistical methods, i.e. g-methods, employed in this project allowed us to analyze the exposure-response data for the DEMS cohort without bias characterized by a time-varying confounder affected by previous exposure, i.e., the HWSE. The use of g-methods also allowed us to evaluate the impacts of hypothetical interventions on exposure on the risks of IHD and lung cancer mortality. Further, once we employed g-methods to control for the HWSE, we found that *more exposure to REC is associated with excess mortality risk from both IHD and lung cancer*.

The MSHA exposure limit for diesel exhaust is 160 $\mu g/m^3$ total carbon (TC) measured on the submicrometer particulate matter fraction. We used internal exposure assessment data from DEMS to convert TC limits to appropriate EC values and subsequently EC values from the submicron level to the appropriate respirable fraction. The equivalent value for 160 $\mu g/m^3$ of TC was 106 $\mu g/m^3$ of REC in DEMS. This is a study specific conversion and the equivalent values of REC with respect to the MSHA TC limit in different settings will not necessarily be the same. TC to REC conversions should rely on study specific data.

The overall findings of this project suggest that the current MSHA exposure limit for diesel exhaust of 160 $\mu g/m^3$ TC is too high to protect miners against excess risk of heart disease and lung cancer. Although MSHA does not have a generic limit for mine dust, there is a 1.5 mg/m^3 exposure limit for respirable coal dust. Our results suggest that MSHA may also consider a standard for generic respirable ore mine dust to reduce the overall heart disease burden

Our lung cancer paper was published in 2016. Later last year the journal *Epidemiology* accepted a commentary on our paper, critical of our methods. Our rejoinder was just accepted and will be published along with the commentary in November, 2017. (See papers in press.) The fact that our paper has already evoked a response will increase its impact.

8.0 Recommendations for Future Work

We have just been funded by the Alpha Foundation to continue to study risk of chronic disease in the DEMS of non-metal miners. Our next focus is on chronic obstructive pulmonary disease (COPD). Like IHD, this chronic lung disease is typically accompanied by symptoms, making it difficult to stay employed in a physically demanding job. As a result, workers who are more affected by shortness of breath, or wheezing, for example, are more likely to leave work prematurely, leading to a HWSE wherein the workers with the healthiest lungs end up accumulating the most exposure. Thus, we will again need to apply analytic methods that can properly control for this bias when estimating the exposure-response between diesel exhaust, RPM, and COPD mortality.

9.0 References:

1. https://arlweb.msha.gov/regs/fedreg/final/2014finl/2014-09084.asp. February 2017.

2. Shah AS, Langrish JP, Nair H, et al. Global association of air pollution and heart failure: a systematic review and meta-analysis. *The Lancet*. 2013;6736(13):1–10.

3. Brook RD, Rajagopalan S, Pope CA, et al. Particulate Matter Air Pollution and Cardiovascular Disease. An Update to the Scientific Statement From the American Heart Association. *Circulation*. 2010;121:2331–2378.

4. Dockery DW, Pope CA, Xu X, et al. An association between air pollution and mortality in six U.S. cities. *The New England journal of medicine*. 1993;329(24):1753–9.

5. Cullen MR. Invited commentary: the search for preventable causes of cardiovascular disease-- whither work? *American journal of epidemiology*. 2009;169(12):1422–5.

6. Attfield MD, Schleiff PL, Lubin JH, et al. The Diesel Exhaust in Miners study: a cohort mortality study with emphasis on lung cancer. *Journal of the National Cancer Institute*. 2012;104(11):869–83.

7. Richardson DB, Loomis D. The impact of exposure categorisation for grouped analyses of cohort data. *Occupational and environmental medicine*. 2004;61(11):930–5.

8. Govindarajulu US, Malloy EJ, Ganguli B, Spiegelman D, Eisen EA. The comparison of alternative smoothing methods for fitting non-linear exposure-response relationships with Cox models in a simulation study. *The international journal of biostatistics*. 2009;5(1):Article 2.

9. Pope CA, Muhlestein JB, May HT, Renlund DG, Anderson JL, Horne BD. Ischemic heart disease events triggered by short-term exposure to fine particulate air pollution. *Circulation*. 2006;114(23):2443–

10. Pope CA, Burnett RT, Thurston GD, et al. Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease. *Circulation*. 2004;109(1):71–7.

11. Kloog I, Koutrakis P, Coull B a., Lee HJ, Schwartz J. Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmospheric Environment*. 2011;45(35):6267–6275.

12. Kawachi I, Colditz GA, Stampfer MJ, et al. Smoking Cessation and Time Course of Decreased Risks of Coronary Heart Disease in Middle-Aged Women. *Arch Intern Med*. 1994;154:169–175.

13. Lightwood JM, Glantz SA. Short-term economic and health benefits of smoking cessation: myocardial infarction and stroke. *Circulation.* 1997;96(4):1089–96.

14. Qiao Q, Tervahauta M, Nissinen A, Tuomilehto J. Mortality from all causes and from coronary heart disease related to smoking and changes in smoking during a 35-year follow-up of middle-aged Finnish men. *European heart journal*. 2000;21(19):1621–6.

15. Applebaum KM, Malloy EJ, Eisen EA. Reducing healthy worker survivor bias by restricting date of hire in a cohort study of Vermont granite workers. *Occupational and environmental medicine*. 2007;64(10):681–7.

16. Robins JM, Hernán MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G, eds. *Longitudinal Data Analysis*. New York, NY: Chapman and Hall/CRC Press; 2009.

17. Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the healthy worker survivor effect. *Math Model.* 1986;7:1393–1512.

18. Robins J. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *J Chronic Dis.* 1987;40(suppl 2):139S–161S.

19. Stewart PA, Coble JB, Vermeulen R, et al. The diesel exhaust in miners study: I. Overview of the exposure assessment process. *Ann Occup Hyg.* 2010;54:728–746.

20. Coble JB, Stewart PA, Vermeulen R, et al. The diesel exhaust in Miners Study: II. Exposure monitoring surveys and development of exposure groups. *Ann Occup Hyg*. 2010;54:747–761.

21. Vermeulen R, Coble JB, Yereb D, et al. The diesel exhaust in Miners Study: III. Interrelations between respirable elemental carbon and gaseous and particulate components of diesel exhaust derived from area sampling in underground non-metal mining facilities. *Ann Occup Hyg*. 2010;54:762–773.

22. Vermeulen R, Coble JB, Lubin JH, et al. The diesel exhaust in Miners Study: IV. Estimating historical exposures to diesel exhaust in underground non-metal mining facilities. *Ann Occup Hyg*. 2010;54:774–788.

23. Stewart PA, Vermeulen R, Coble JB, et al. The diesel exhaust in miners study: V. Evaluation of the exposure assessment methods. *Ann Occup Hyg*. 2012;56:389–400.

24. Yanowitz J, McCormick RL, Graboski MS. In-use emissions from heavy-duty vehicle emissions. Environ Sci Technol. 2000;34: 729–40.

25. Silverman DT, Samanic CM, Lubin JH, et al. The diesel exhaust in Miners study: a nested case-control study of lung cancer and diesel exhaust. *J Natl Cancer Inst*. 2012;104:855–868.

26. Richardson D, Wing S, Steenland K, McKelvey W. Time-related aspects of the healthy worker survivor effect. *Annals of epidemiology*. 2004;14(9):633–9.

27. MSHA. Diesel Particulate Matter Exposure of Underground Metal and Nonmetal Miners, Final Rule. 30 CFR Part 57.5060 U.S. Mine Safety and Health Administration; 2001.

28. ACGIH Diesel exhaust (particulate and particulate adsorbed components), draft TLV-TWA document. Cincinnati, OH: American Conference of Governmental InRPMrial Hygienists; 2001.

29. Taubman SL, Robins JM, Mittleman MA, Hernán MA. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *Int J Epidemiol*. 2009;38:1599–1611.

30. Steenland K, Armstrong B. An Overview of Methods for Calculating the Burden of Disease Due to Specific Risk Factors. *Epidemiology*. 2006;17(5):512–519.

33. Vermeulen R, Silverman DT, Garshick E, Vlaanderen J, Portengen L, Steenland K. Exposure-response estimates for diesel engine exhaust and lung cancer mortality based on data from three occupational cohorts. *Environ Health Perspect*. 2014;122:172–177.

10.0 Appendices: None.

11.0 Acknowledgement/Disclaimer

This study was sponsored by the Alpha Foundation for the Improvement of Mine Safety and Health, Inc. (ALPHA FOUNDATION). The views, opinions and recommendations expressed herein are solely those of the authors and do not imply any endorsement by the ALPHA FOUNDATION, its Directors and staff.