Original Article

Occupational diesel exposure, duration of employment, and cumulative risk of lung cancer: an application of the parametric g-formula.

Andreas M Neophytou[a]; Sally Picciotto[a]; Sadie Costello[a]; Ellen A Eisen[a]

[a]Division of Environmental Health Sciences, UC Berkeley School of Public Health, Berkeley, CA, USA

Corresponding author:

Dr. Andreas M. Neophytou

50 University Hall #7360

Berkeley, CA 94720-7360

Email: aneophytou@berkeley.edu

Phone number: (510) 643-4613

Running Head: Diesel exposure interventions and lung cancer risk

Abstract

Background: If less healthy workers terminate employment earlier, thus accumulating less exposure, yet remain at greater risk of the health outcome, estimated health effects of cumulative exposure will be biased downward. If exposure also affects termination of employment, then the bias cannot be addressed using conventional methods. We examined these conditions as a prelude to a reanalysis of lung cancer mortality in the Diesel Exhaust in Miners Study designed to evaluate the effects of hypothetical interventions.

Methods: We applied an accelerated failure time model to assess the effect of exposures to respirable elemental carbon (REC, surrogate for diesel) on time to termination of employment in the subcohort of non-metal miners who ever worked underground (n=8,307). We then applied the parametric g-formula to assess how possible interventions setting REC exposure limits would have changed lifetime risk of lung cancer, adjusting for time-varying employment status.

Results: The median time to termination was 36% shorter (95% confidence interval (CI): 33% – 39%), per IQR increase in REC exposure. Lung cancer risk decreased when interventions became more stringent, with a risk ratio of 0.79 (95% CI: 0.45 – 1.13) comparing a limit of ≤25 $\mu g/m^3$ REC to no intervention. The fraction of cases attributable to REC was 27% in this population.

Conclusions: The g-formula controlled for time-varying confounding by employment status that was also affected by diesel exposure, the signature of healthy worker survivor bias. It also offers an alternative approach to risk assessment for estimating excess cumulative risk, and the impact of interventions based entirely on an observed population.

1

Adverse health effects, including lung cancer, have been associated with occupational exposure to diesel exhaust.[1-5] Diesel exhaust is a mixture of elemental carbon (EC), polycyclic aromatic hydrocarbons (PAHs), condensed engine oil, other organic compounds, and exhaust gases. EC is the component frequently used as a primary surrogate for diesel exhaust exposure.[6-8] High exposures to diesel exhaust due to widespread use of diesel-powered engines have been common in industries such as trucking, mining, and railroads. Based in large part on exposure-response estimates for EC exposure and lung cancer risk from occupational studies of cohorts of workers employed in these industries[9-12], the International Agency for Research on Cancer (IARC) has classified diesel exhaust as a human carcinogen.[13,14]

Previous studies have not accounted for the healthy worker survivor bias using methods that address time-varying confounding affected by previous exposure. Healthy worker survivor bias 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 will not be 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, then the effect of diesel exhaust on lung cancer may be even higher than previously reported.

We begin this reanalysis of the Diesel Exhaust in Miners Study (DEMS) by evaluating the relationship between respirable EC (REC) exposure and time to termination of employment, a key potential component of healthy worker survivor bias. To accomplish this we use an accelerated failure time model, a method particularly suited for time to event outcomes. We then use the parametric g-formula to assess the impacts of hypothetical REC exposure limits on the risk of lung cancer, controlling for work status. The parametric g-formula is one of a broader

group of 'g-methods'[15] first introduced by Robins[16,17] to control for time-varying confounding affected by past exposure, a signature characteristic of healthy worker survivor bias in longitudinal occupational studies.[18] The parametric g-formula has been applied in several longitudinal studies,[19,20] but only recently in an occupational study of asbestos exposure and risk of lung cancer.[21,22] It has also been used to assess occupational radon exposure interventions and cumulative risk of lung cancer.[23]

In the current study we assess the possible impact of hypothetical interventions setting maximum levels of exposure to REC in the DEMS ever-underground miners subcohort. To identify a range of relevant exposure limits, we considered a recent meta-regression analysis,[24] as well as current and past exposure limits for EC, either as enforced by the Mine Safety and Health Administration (MSHA)[25] or recommended by the American Conference of Governmental Industrial Hygienists (ACGIH).[26] We also estimate the attributable fraction of lung cancers due to diesel exhaust exposure. This application assesses effects of different exposure limits on adverse health effects within a public health framework by simulating what would have happened if various exposure limits had been enforced in the whole sub-cohort.

METHODS

Population

The DEMS cohort is described in detail elsewhere.[11] Briefly, eight non-metal mining facilities were selected to participate in the study. Facilities were chosen so that participants would have low levels of potentially confounding exposures such as radon, asbestos and silica; there was extensive use of diesel power engines, and sufficient time since dieselization for lung cancer development. Workers employed in a blue collar job at a mine for at least one year after

dieselization were eligible. The final sample size of the DEMS cohort was 12,315 workers, and follow-up began at each mine after dieselization, ranging from 1947 to 1967. The analysis in the current study was restricted to workers who ever worked underground (n=8,307, "ever-underground"), as opposed to workers working only on the surface, using definitions proposed by Attfield et al.[11] and Silverman et al.[12] based on job location. Exposure differed substantially between the two groups, because of differing sources and ranges of diesel exposure.[27-29] The DEMS data used in this study was a publicly available dataset obtained under a data use agreement with the National Institute of Occupational Safety and Health (NIOSH) and the National Cancer Institute (NCI).

Outcome and Covariates

Work records provided information on job titles and dates as well as sex, date of birth, and partial information on race. Vital status was ascertained via matching with the National Death Index (NDI-plus) and the Social Security Administration (SSA) death files. The 111 participants who could not be matched to NDI-Plus or SSA were considered alive until their last observed date and censored afterwards. Cause of death was obtained from NDI-plus from 1979 to 1997 and prior to that from death certificates coded by a certified nosologist. In the ever-underground subcohort, 122 lung-cancer deaths were ascertained. Table 1 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 or Hispanic. The median year at start of follow up was 1975 and the mean age (± SD) at baseline was 30.23 (±9.05). The age at death due to lung cancer ranged from 44 to 90 years of age.

Exposure

Exposure assessment in DEMS has been described in detail elsewhere.[27-31] 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, these effects were estimated using historical CO measurements under the assumption that global changes in these determinants would have affected CO and REC similarly. Annual average daily exposure was assigned to each year of active employment for each participant during follow-up. The mean (SD) annual average daily 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 (IQR) of 167.3 $\mu g/m^3$. Exposure assessment was performed blind to outcomes.

Statistical Methods

Termination of Employment

In order to evaluate whether employment termination is affected by the exposure, in which case traditional regression will give biased results, we fit an accelerated failure time model to estimate the effect of REC exposure on time to termination of active employment. We assume an exponential distribution of time to termination as evidenced in Figure 1.

For this model, the data were truncated to only include person time with active work history. All last observed dates of active work history prior to December 31, 1997 for those still alive were considered terminations of active employment. 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

5

daily exposure as a continuous variable, as well as sex, race, facility, birth year and a natural cubic spline for calendar year. The process was repeated restricting to incident hires, where follow-up time is 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 1, the inverse of the time to termination ratios obtained by the accelerated failure time model can be interpreted as hazard ratios 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, state, and cumulative exposure, with attained age as the time scale.

Interventions

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 assess 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 μg/m³ for everyone), and four different maximum annual average daily REC exposures. The four exposure limits evaluated were based on the past and current MSHA limits of 400 and 160 μg/m³ of total carbon exposure,[25] and the past and current ACGIH recommendations of 50 and 20 μg/m³ EC exposure,[26] all measured on the submicrometer particulate matter fraction. We used internal exposure assessment data[29] to convert total carbon limits to appropriate EC values and subsequently EC values from the submicron level to the appropriate respirable fraction. The corresponding study-specific intervention values based on past and current MSHA limits were

260 and 106 µg/m³ REC, and those based on ACGIH recommendations were 65 and 25 µg/m³ REC.

The parametric g-formula

Under the assumptions of (a) no unmeasured confounders (conditional exchangeability) at all time points, (b) counterfactual consistency i.e., every individual's counterfactual outcome under their observed exposure history is equal to their observed outcome, and (c) correct model specification., the parametric g-formula is a generalization of standardization for time-varying exposures and covariates.[16,19,32,33] 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 is described in detail in the Appendix. Briefly, it 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, state, and indicator for workers hired prior to dieselization). Person-time used in these models was restricted to age less than or equal to 90. This age cut-off was chosen because there were no cases after that age and follow-up was long enough to allow estimates of cumulative risk

to be interpreted as lifetime risk. Both REC exposure and employment status were lagged by 15 years in outcome models to account for cancer latency, and current annual average daily REC exposure values as well as cumulative exposure up to the previous year were included in outcome models 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. We preformed the same process with a larger pseudo-sample (n=50,000) as a sensitivity analysis. Exposure and covariate values at each age 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. For example, under the intervention of maximum REC exposures of 65 $\mu g/m^3$, all predicted values above 65 $\mu g/m^3$ were replaced with 65 $\mu g/m^3$ and otherwise remained unchanged. 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.

The cumulative incidence of lung cancer, with age as the time scale, was then calculated for each intervention using a cumulative incidence estimator for the sub-distribution of the event of interest, in the presence of competing risks.[34] Confidence Intervals (CI) for cumulative incidences and risk ratios (RR) were obtained 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) based on the gformula SAS macro available at: http://www.hsph.harvard.edu/causal. For further details on model specifications and the simulation process refer to the Appendix and to Taubman et al.[19] or Cole et al.[22]

RESULTS

Hazard Ratios and 95% CI for an IQR increase in REC exposure and termination of employment are summarized in Table 2. An IQR 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 restricting incident hires to those with at least 3 and 5-year tenures, but remained significant.

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 (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 (Figure 2). Figure 3 shows the simulated cumulative incidence under all proposed interventions; Risk Ratios (RR) comparing the risk under each intervention to the simulated risk under no intervention, are presented in Table 3. The lifetime cumulative incidence when everyone was always unexposed

to REC was estimated at 5.7%, for a risk difference of 2.1% compared to the observed risk of 7.8%. The attributable fraction (AF=(RR-1)/RR)[35] 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.

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 1 lung cancer death per 1000 workers. Similar to conventional risk management analyses, we assumed that all workers were continuously exposed between ages 20-65. Using the observed background rates in the study population, we predict that REC exposure would need to be limited to below 1 $\mu g/m^3$ in order to achieve this level of excess risk.

DISCUSSION

Recent observational epidemiology studies, including a meta-analysis, have found excess risk of lung cancer in relation to exposure to diesel exhaust[9-12,24]. Substantial attention has been paid to whether or not these studies have over stated the risks of diesel exhaust, in particular the DEMS study. In this paper, we examine a factor that could lead to an underestimate of risk of lung cancer, the healthy worker survivor bias, and apply the parametric g-formula in the DEMS subcohort of 8,307 ever-underground miners. We estimated lifetime risks for 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 result in a downward bias in the effect of

cumulative exposure on the outcome, because workers more susceptible to health outcomes may leave work and accumulate less exposure. The effect was attenuated among those with at least 5 years of tenure, consistent with previous results reported by Attfield et al. showing strong dose-response after restricting to those with at least 5-years tenure.[11] 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.

Treating termination of employment as a survival outcome, rather than a binomial one, has several advantages. First, the accelerated failure time model allows us to assess the effect of potential exposures on time to event rather than on the probability of the event. Everyone in an occupational cohort with long enough follow-up time will eventually terminate employment, and time to event is a more useful metric for events that are inevitable given sufficient follow-up time, as is also generally true for all-cause mortality. A limitation of this analysis was the fact that termination of work histories did not necessarily mean retirement. We had no information on whether or not participants found other employment after terminating work at the study facilities. Nevertheless, the purpose of this analysis was to see if exposure predicts active employment status within the participating facilities, as this is the covariate predicting future exposure in the context of this study.

We implemented the parametric g-formula to assess the effect of interventions within a public health framework in a real world setting. The parametric g-formula allows us to estimate effects of such hypothetical interventions using observed data. Additionally, 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, it allows us to adequately control for time-varying confounding affected by prior exposure. The g-formula can

be used to obtain estimates of cumulative risk, risk ratios and risk differences, as well as conditional survival probabilities, which are more easily interpretable with respect to specific population effects and less prone to the limitations of the Hazard Ratio.[19,36] Compared to other g-methods, it also offers flexibility in modelling the exposure and covariates as well as allowing for joint and dynamic interventions, which would be more cumbersome under g-estimation. It is also less prone to bias in the presence of violations of positivity, compared to methods utilizing inverse probability weights.[17,33] Previous studies have successfully used causal inference methods to estimate measures of public health impact in occupational epidemiology, relying primarily on estimation of structural nested failure time models[37,38]. Interpretation of those results, however, becomes somewhat complex in the presence of competing risks, whereas the g-formula is better equipped to handle competing risks as it allows the estimation of sub-distribution functions in the presence of competing risks[39].

One major limitation of the parametric g-formula is the 'g-null paradox', which states that under the causal null hypothesis it may be impossible to correctly specify the models for the g-formula, and as a consequence the causal null hypothesis will be rejected when it is in fact true.[18] Given the weight of the previous evidence for diesel exposure and lung cancer,[9-12,24] however, we believe that the causal null is likely false in this case and therefore none of the interventions in question is likely to suffer from the 'g-null paradox'. A more serious limitation of our analysis is the number of models required, and the parametric assumptions related to each one. These models may propagate biases over a long follow-up period if assumptions of no unmeasured confounding, correct model specification and/or no information bias are violated in one or more models.[19]

A limitation of the DEMS cohort was the lack of data on smoking, a major predictor of lung cancer. A nested case-control study within the DEMS cohort did, however, collect data on smoking, and reported an inverse relationship between smoking and diesel exhaust exposure among ever-underground workers, leading to an overall negative confounding effect for the relationship between cumulative diesel exhaust exposure and lung cancer mortality in the ever-underground subcohort. Smoking adjusted HRs for ever-underground workers from the case-control study were higher than unadjusted HRs from the cohort study, potentially due to downward confounding by smoking[12]. Effect estimates based on cumulative risks under interventions reported in this study are subject to this potential limitation. Absence of smoking data, may have also limited our ability to properly account for competing risks as it is also a major predictor of other causes of mortality.

The current study estimated the effect of interventions to decrease diesel exhaust, represented by REC in this study, on the lifetime lung cancer risk of a real-world occupational cohort of miners with dynamic work histories. These interventions are derived from existing occupational standards and guidelines expressed as limits on exposure as measured by personal samples of an exposed miner during a full shift. Interventions based on the values proposed should lower REC levels 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 standard risk management practice, the goal is to identify exposure limits resulting in risks that do not exceed benchmarks deemed as 'acceptable risk'. The Occupational Safety and

Health Administration (OSHA) has typically relied on less restrictive risk ranges compared to other regulatory agencies such as the EPA, taking technological and economic feasibility within industry into account. A risk of 1 in 1000 has been identified as 'significant risk' by the U.S Supreme Court[40] and has been used by OSHA as a possible starting goal. In this study, we found that an exposure of less than 1 µg/m^3 REC would be needed to achieve this level of lung cancer risk. Despite inherent uncertainties in this number, it suggests that a health based exposure limit would make it difficult to use diesel equipment underground.

The observed excess risk of lung cancer associated with diesel exhaust is qualitatively consistent with the existing evidence. The intervention effects, however, reported here as quantitative estimates of cumulative risks or risk reductions, may not be generalizable. The estimated risk reductions used the study population observed risk in the study population as the point of reference, and are a function of both the particular exposure response between diesel exhaust and lung cancer and also the number of people affected by each specified intervention. The number of subjects above each exposure limit, in turn, is a function of the exposure distribution in the current study. The small reduction in risk we observed when the REC limit was set at <=260 µg/m^3, is mostly due to the limited person time above exposure cut-off, i.e. the intervention would have left the majority of the cohort unaffected. The same intervention in a different population with the same exposure-response profile and a large amount of person-time exposed above 260 µg/m^3 would have resulted in a larger reduction in risk. Despite limiting generalizability, our use of the unexposed age specific lung cancer rate estimated directly from this study, rather than the background rate in the general population, is also an advantage. It avoids the questionable assumption that mortality rates in the general population are

transportable to occupational cohorts, implicit in life table calculations for occupational risk assessment.

Under alternative intervention assumptions 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 EC, (Silverman et al.[12] and Vermeulen et al.[24]), which found excess lung cancer risk in the lower observed ranges (1-25 $\mu g/m^3$) of occupational EC exposures.

CONCLUSION

IARC has classified diesel exhaust as a human carcinogen, and evidence from large occupational studies show an excess risk of lung cancer associated with a wide range of exposures.[13,14,25] We found an increased risk of lung cancer and also a shorter time to termination of employment associated with higher diesel exhaust exposure, suggesting presence of healthy worker survivor bias. Exposure limits lower than the ones currently enforced by MSHA could further reduce the lifetime risk of lung cancer among workers in the mining industry exposed to diesel exhaust.

REFERENCES

1. Diesel Working Group. Diesel Exhaust: A Critical Analysis of Emissions, Exposure, and Health Effects. Special Report. Cambridge, MA:Health Effects Institute; 1995.

2. Bhatia R, Lopipero P, Smith AH. Diesel exhaust exposure and lung cancer. *Epidemiology*. 1998;9:84–91.

3. Steenland NK, Silverman DT, Hornung RW. Case–control study of lung cancer and truck driving in the Teamsters Union. *Am J Pub Health*. 1990;80:670–674.

4. Garshick E, Laden F, Hart JE, et al. Lung cancer in railroad workers exposed to diesel exhaust. *Environ Health Perspect*. 2004;112:1539–1543.

5. Garshick E, Laden F, Hart JE, et al. Lung cancer and vehicle exhaust in trucking industry workers. *Environ Health Perspect*. 2008;116:1327–1332.

6. Birch ME, Cary RA. Elemental carbon-based method for occupational monitoring of particulate diesel exhaust: methodology and exposure issues. *Analyst*. 1996;121:1183–1190.

7. Birch ME, Noll JD. Submicrometer elemental carbon as a selective measure of diesel particulate matter in coal mines. *J Environ Monit*. 2004;6:799-806.

8. Bunn WB III, Valberg PA, Slavin TJ, Lapin CA. What is new in diesel. *Int Arch Occup Environ Health*. 2002;75(suppl):S122-S132.

9. Steenland K, Deddens J, Stayner L. Diesel exhaust and lung cancer in the trucking industry: exposure–response analyses and risk assessment. *Am J Ind Med*. 1998;34:220–228.

10. Garshick E, Laden F, Hart JE, Davis ME, Eisen EA, Smith TJ. Lung cancer and elemental carbon exposure in trucking industry workers. *Environ Health Perspect*. 2012;116:1301–1306.

11. Attfield MD, Schleiff PL, Lubin JH, et al. The Diesel Exhaust in Miners Study: a cohort mortality study with emphasis on lung cancer. *J Natl Cancer Inst*. 2012;104:869–883.

12. 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(11):855–868.

13. Benbrahim-Tallaa L, Baan RA, Grosse Y, et al. Carcinogenicity of diesel-engine and gasoline-engine exhausts and some nitroarenes. *Lancet Oncol*. 2012;13:663–664.

14. IARC. IARC Monographs on the evaluation of the carcinogenic risks to humans. Diesel and Gasoline Engine Exhausts and Some Nitroarenes Volume 105. Lyon, France: International Agency for Research in Cancer; 2013.

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

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

17. 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.

18. Eisen EA, Robins JM, Picciotto S. Healthy Worker Effect. In: El-Shaarawi AH, Piegorsch W, (eds). *Encyclopedia of Environmetrics,* 2nd ed. Chichester, UK: John Wiley & Sons Ltd; 2012:1269-1272.

19. 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*. 2008;38:1599-1611.

20. Young JG, Cain LE, Robins JM, O'Reilly EJ, Hernán MA. Comparative effectiveness of dynamic treatment regimes: an application of the parametric G-formula. *Stat Biosci*. 2011;3:119-143.

21. Naimi AI, Cole SR, Hudgens MG, Brookhart MA, Richardson DB. Assessing the component associations of the healthy worker survivor bias: occupational asbestos exposure and lung cancer mortality. *Ann Epidemiol*. 2013;23:334-341.

22. Cole SR, Richardson DB, Chu H, Naimi AI. Analysis of occupational asbestos exposure and lung cancer mortality using the g-formula. *Am J Epidemiol*. 2013;177(9):989-996.

23. Edwards JK, McGrath LJ, Buckley JP, Schubauer-Berigan MK, Cole SR, Richardson DB. Occupational radon exposure and lung cancer mortality: estimating intervention effects using the parametric g-formula. *Epidemiology*. 2014;25(6): 829-834.

24. 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.

25. 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.

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

27. 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.

28. 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.

29. 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.

30. 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.

31. 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.

32. Robins JM. Causal inference from longitudinal data. In: Berkane M, (ed). *Latent Variable Modeling and Applications to Causality. Lecture notes in statistics.* Vol. 120. Springer-Verlag; 1997:69-117.

33. Robins JM, Hernán MA, Siebert U. Effects of multiple interventions. In:Ezzati M, Lopez AD, Rodgers A, Murray CJL, (eds). *Comparative Quantification of Health Risks: Global*

and Regional Burden of Disease Attributable to Selected Major Risk Factors. Geneva: World Health Organization; 2004.

34. Lau B, Cole SR, Gange SJ. Competing Risk Regression Models for Epidemiologic Data. *Am J Epidemiol.* 2009;170(2): 244-256.

35. 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.

36. Hernán, MA. The hazards of hazard ratios. *Epidemiology.* 2010;21:13-15.

37. Picciotto S, Chevrier J, Balmes J, et al.. Hypothetical interventions to limit metalworking fluid exposures and their effects on COPD mortality: G-estimation within a public health framework. *Epidemiology.* 2014;25(3): 436--443.

38. Picciotto S, Peters A, and Eisen E. Hypothetical exposure limits for straight metalworking fluids and cardiovascular mortality in a cohort of autoworkers: structural accelerated failure time models in a public health framework. *Am J Epidemiol.* 2015;181(8):563-570.

39. Naimi AI and Tchetgen Tchetgen EJ. Estimating Population Impact in the Presence of Competing Events. *Am J Epidemiol.* 2015;181(8):571-574.

40. Industrial Union Department, AFL-CIO v. American Petroleum Institute, 448 U.S. 601, 65L. Ed. 2D 1010, 100 S. Ct. 2844 (1980).

Figure 1: Time to termination of employment curve in the DEMS ever-underground subcohort (solid line), compared to a parametric exponential survival function with hazard λ=0.06 (dashed line).

Figure 2: Observed (solid line) and simulated (dashed line) cumulative lung cancer mortality in the DEMS ever-underground cohort, under the natural course.

Figure 3: 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.

Table1

Table 1: Demographic characteristics at baseline in the DEMS ever-underground cohort

Variable	
No. of participants (No)	8307
Male, No. (%)	7478 (97)
Race, No. (%)	
Unknown	3773 (49)
White/Hispanic	3776 (49)
Black	127 (2)
Mine type, No. (%)	
Limestone	946 (11)
Potash	3278 (40)
Salt	497 (6)
Trona	3586 (43)
Age at start of follow-up, mean (sd)	30.23 (9.04)
Duration of active work history in years, mean (sd)	20.04 (9.53)
Lung cancer deaths (No)	122
Total deaths (No)	1404

Table2

Table 2: Hazard Ratio approximations with 95% CIs from conditional accelerated failure time models for an IQR (=167 µg/m³) 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)	1.57 (1.47 – 1.62)
Incident hires[b] only (n=7,750)	1.54 (1.44 – 1.60)
Incident hires with ≥3 years tenure (n=5,993)	1.33 (1.26 – 1.40)
Incident hires with ≥5 years tenure (n=4,971)	1.24 (1.17 – 1.30)

[a]This approximates the HR for the event comparing always exposed above and below the cut-off 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

Table3

Table 3: Cumulative incidence for lung cancer mortality and corresponding Risk Ratios (RR) and 95% CIs, under several interventions in the DEMS ever-underground cohort.

Intervention[a]	Cumulative Incidence	RR (95% CI)	% Person-time above Intervention limit	% Active work person-time above Intervention limit
Observed Data	7.79	.	.	.
[a]Simulated Natural Course	7.80	.	.	.
[b]Max REC Limit at ≤260 μg/m^3	7.73	0.99 (0.92 – 1.04)	5	10
[b]Max REC Limit at ≤106 μg/m^3	7.14	0.91 (0.65 – 1.17)	20	39
[b]Max REC Limit at ≤65 μg/m^3	6.80	0.87 (0.56 – 1.17)	28	53
[b]Max REC Limit at ≤25 μg/m^3	6.18	0.79 (0.45 – 1.13)	33	63
[c]Always unexposed	5.71	0.73 (0.38 – 1.09)	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
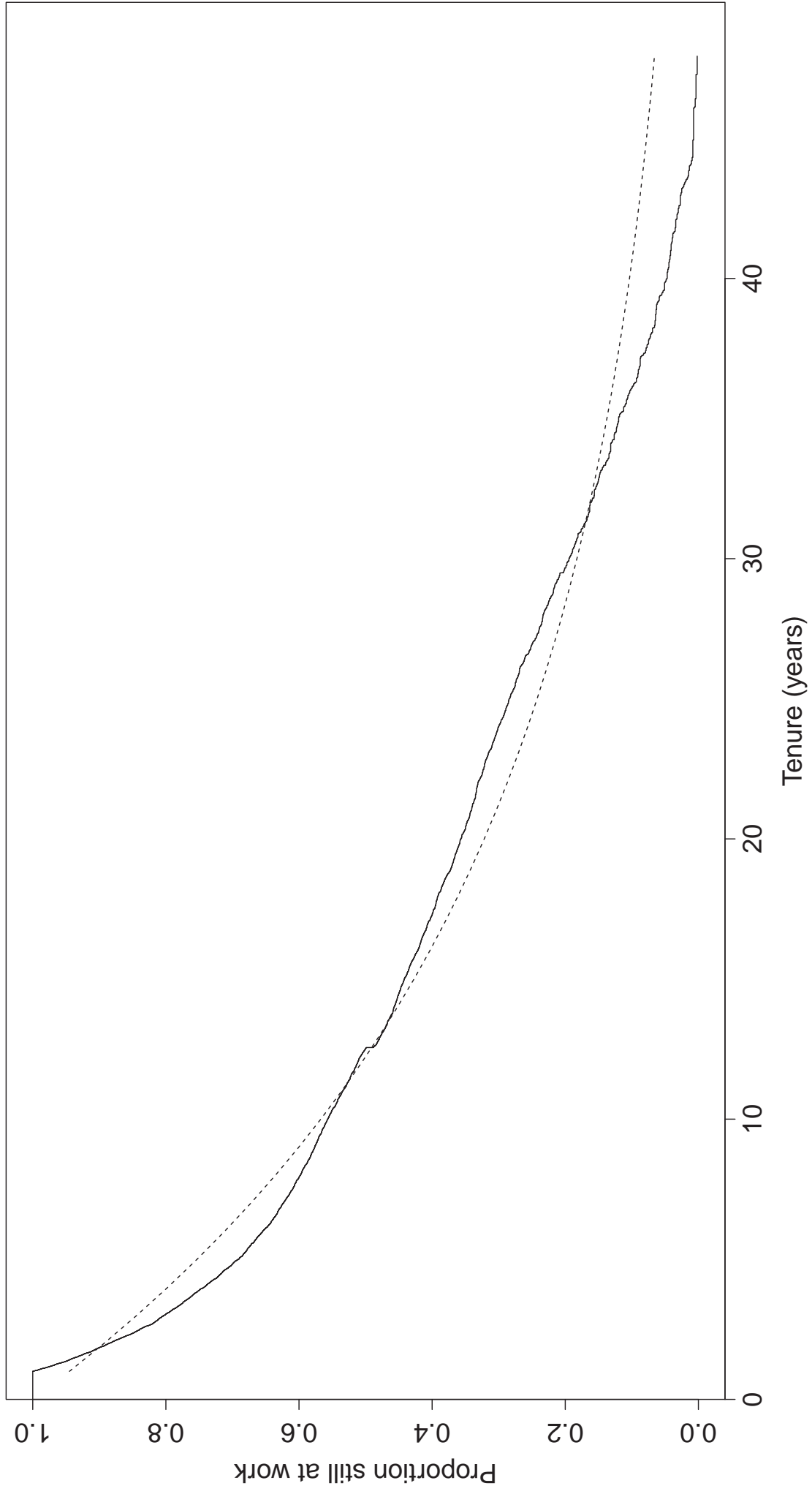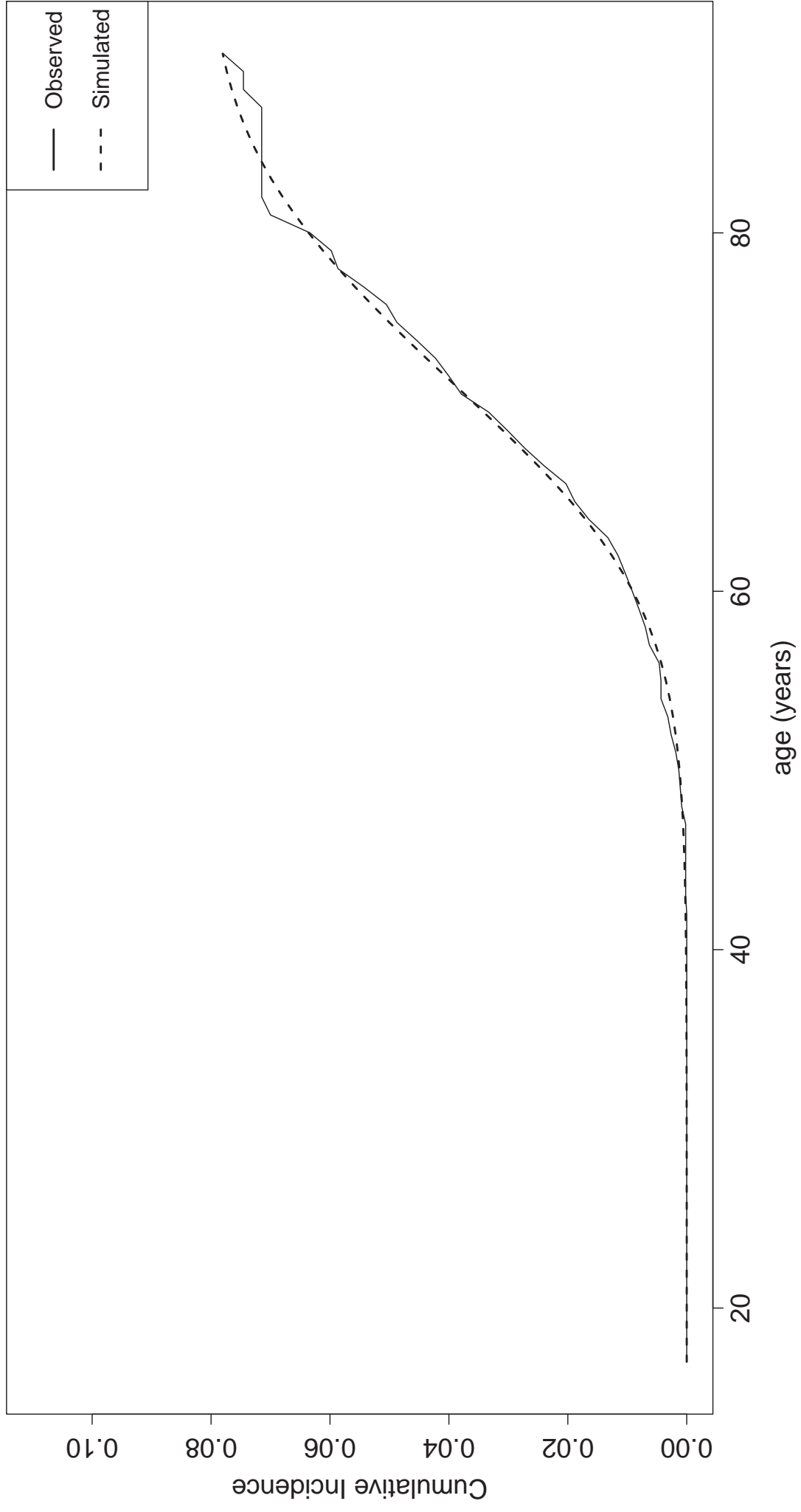[c]Under always unexposed REC is set to 0 at all times.

Figure1

Figure2

Figure3