Healthy worker survivor bias in the Colorado Plateau uranium miners cohort

Alexander P Keil, David B Richardson, and Melissa A. Troester

Running Head: Healthy worker survivor bias in uranium miners

Correspondence to Dr. Alexander P. Keil, Department of Epidemiology, University of North Carolina, 2102E McGavran-Greenberg Hall, Campus Box 7435, Chapel Hill, NC 27599-7435, akeil@unc.edu, t: (408)768-3960, f: (919) 966-2089

ABSTRACT

Cohort mortality studies of underground miners have been used to estimate the number of lung cancer deaths attributable to radon exposure. However, previous radon-lung cancer association studies among underground miners may be subject to healthy worker survivor bias, a type of time-varying confounding by employment status. We examined radon-mortality associations in a study of 4124 male uranium miners from the Colorado Plateau followed from 1950 through 2005. We estimated the time ratio (the relative change in the median survival time) per 100 working level months (radon exposure averaging 130,000 mega-electron volts of potential alpha energy per liter, per working month) using g-estimation of structural nested models. After controlling for healthy worker survivor bias, the time ratio (95% confidence intervals) for lung cancer per 100 working level months was 1.168 (1.152, 1.174). In an unadjusted model the estimate was 1.102 (1.099, 1.112), 39% lower. Controlling for this bias, we estimated that among 617 lung cancer deaths, 6,071 person-years of life were lost due to occupational radon exposure during follow-up. Our analysis suggests healthy worker survivor bias in miner cohort studies can be substantial, warranting re-examination of current estimates of radon's estimated impact on lung cancer mortality.

Keywords: cohort, dose-response, g-estimation, lung neoplasms, mortality, occupational, radon, structural nested model

Abbreviations

HR: Hazard Ratio
SNAFT: Structural nested accelerated failure time
TR: Time ratio
WLM: Working level month

Radon is a ubiquitous gas that concentrates in indoor air and is a leading cause of lung cancer in the United States. The burden of lung cancer attributable to residential radon exposure is of considerable interest, given costs of compliance with the current United States Environmental Protection Agency action level and public health impacts of a lower action level recommended by the President's Cancer Panel (1, 2, 3). Radon exposure is protracted and may have persistent effects, so researchers frequently model radon-lung cancer associations using a cumulative metric of radon exposure. Occupational cohort mortality studies of underground miners have contributed to risk assessment, providing influential estimates of radon-lung cancer associations (3, 4, 5, 6, 7). Occupational studies are better suited than residential studies for estimating precise dose-response parameters and exploring time-related aspects of the exposure-outcome association because variation in long-term exposures is better characterized and occupational exposures often reflect a broader dose range (5). However, occupational studies are subject to unique biases that impact their utility for characterization of dose-response functions.

One of the biases particular to occupational settings is healthy worker survivor bias. This bias results when workers at higher risk for the outcome of interest tend to leave work at higher rates than workers at lower risk. When the exposure of interest is aggregated over time this phenomenon can result in higher exposures among healthier individuals (8). Consequently, disease rates of employed and unemployed individuals are generally not comparable, even among those with identical cumulative exposure. Thus, healthy worker survivor bias can be conceptualized as a form of confounding by employment status (9).

Regression methods can be used to control confounding by employment status in some cases, and they are typically used to estimate exposure-response metrics for radon, often stratified by proxies of employment status, such as employment duration (3, 5, 7, 10, 11). However, regression methods cannot completely control this bias when exposure in the past affects subsequent employment (Figure 1) (12). The potential for this bias has not been evaluated in miner studies, despite the availability of relevant methodologic advances (13, 14). Herein, we apply methods for controlling confounders affected by past exposure to estimate dose-response parameters between radon and lung cancer and all-cause mortality.

MATERIALS AND METHODS

Study population

The Colorado Plateau uranium miners' cohort includes 4,137 miners who agreed to participate in a health study by the United States Public Health Service, completed at least one health exam and interview between January 1, 1950 and December 31, 1959 and who were currently mining or started mining during follow-up (15).

Follow-up for mortality was assessed through 12/31/2005 (16, 17). Cause of death information was obtained directly from death certificates before 1979 and the National Death Index for deaths occurring thereafter. We define death from lung cancer using the code for the underlying cause of death indicating malignant neoplasms of the trachea, bronchus or lung (using ICD revision in use at time of death).

Monthly radon exposures in working level months (WLM – defined as any combination of exposure rate in working levels [130,000 mega-electron volts of potential alpha energy per liter of air]) were derived from raw data files (17). These exposure data were originally derived from a job-exposure matrix using area measurements and extrapolations from nearby mineshafts, mines, or regional averages. Estimated radon exposure due to previous work in hard-rock (i.e. non-uranium) mines was also recorded. Three miners were excluded who had lifetime cumulative exposures greater than 10,000 WLMs.

Individual information on smoking histories was obtained from surveys conducted in 1985 or from prior surveys (for decedents or non-respondents). We excluded 10 miners with unknown smoking status. Employment status (active versus inactive) was assumed to be continuous between hire and termination dates.

Our analytic dataset included a record for every person-month between study enrollment and the earliest of death, loss to follow-up, or 12/31/2005.

Statistical Methods

We use an accelerated failure time model to estimate the change in the expected age at death due to an increment of cumulative radon exposure under a linear dose-response assumption. This quantity is expressed as the time ratio (TR) and is reported along with associated 95% confidence intervals for a 100 WLM increase in cumulative radon exposure. With respect to time varying cumulative exposures, the TR can be interpreted as the relative change in the median remaining survival time after a unit increase in the exposure of interest. For example, if an individual would survive to age 70 in the absence of exposure but only to age 60 if exposed at age 20, then the time ratio for a unit increase in cumulative exposure is given as (70-20)/(60-20) = 1.25.

Inference in accelerated failure time models is similar to that in models for the hazard ratios or disease rate ratios. Under an exponential survival time distribution, the TR (transformed so that a TR > 1 indicates harmful exposure) and hazard ratio will be identical, though this equivalence does not hold for other distributions (18). Our exposure of interest was the radon exposure that accumulates after study enrollment, and we defined employment history as the cumulative time at work after enrollment. We estimate TRs for lung cancer mortality and all-cause mortality.

We estimated TRs using a structural nested accelerated failure time model (SNAFT model) fit by g-estimation (13). Here we provide a basic explanation of the SNAFT model in a study in which the age at death is known for all individuals. In Web Appendix 2, we fully describe our approach with the miner data, in which some of the deaths are censored.

We used age as the analytic time scale, and we defined entry into the study as the age at first health exam. Some entry exams were conducted long after hire because uranium mining in the Colorado plateau began before 1950. This may be problematic because any deaths before 1950 would not be recorded, leading to study entry criteria that depended on remaining alive and employed. Robins (9, pg. 1435) refers to this process as "selection bias by cohort definition," which is not addressed by treating employment status as a time varying confounder. Following Robins, we considered exposure estimates and employment duration before study entry to be time-fixed covariates. Other approaches are considered below. Cumulative exposure and employment duration was defined as zero at entry. Cumulative radon exposure began accruing only after a 5-year lag from the study entry, while employment status was not lagged.

Our SNAFT model was:

$$T^0 = m + \int_m^T (1 + \phi \bar{X}_{k-60}) \; dk \qquad (1)$$

Where T is the observed age at death, in months, m is the age at study entry, $\bar{X}_{k-60}$ is cumulative radon exposure with a 60-month (5-year) lag, ϕ is the parameter of interest and T^0 is the survival time that would be expected under no radon exposure during follow-up. Time is denoted by k. T^0 is an individual level variable that can be deterministically derived from the model shown in (1), a value of ϕ and the observed quantities: age at death, cumulative exposure, and age at entry.

Consistent with much of the prior radon literature, in which the excess relative rate (rate ratio – 1) is modeled on a linear scale (19), the parameter ϕ is defined as the excess relative time (where TR $= 1 + \phi$). Our novel approach contrasts with previous uses of SNAFT models, which are typically log-linear (e.g. Hernan et al (2005)(20)). In contrast with a log-linear model, our model is a linear, rather than multiplicative, model for the time ratio. As a technical note, our model places no bounds on ϕ and thus does not exclude negative values increments of the baseline survival time (the integrand term in model 1) in the case of beneficial exposures.

Consequently, use of our model is best suited to associations between health outcomes and agents with known deleterious effects, such as radon. As long as ϕ multiplied by the maximum observed exposure is less than one, this condition will not bias the estimate of ϕ. Thus, studies in which exposures are low (as in residential studies of radon) may not be subject to this caveat even when some studies may be expected to yield estimates below the null by sampling variability.

In SNAFT models, the baseline time, T^0, can be interpreted as a potential outcome representing the time of death we would observe, had we intervened to prevent exposure at work (for example, by mandating the use of 100% efficient respirators). This interpretation allows one to easily calculate the years of life lost (among cases) due to occupational radon exposure as $T - T^0$, which we use to supplement the TR as an estimate of the impact of radon exposure (21). We calculated the years of life lost due to exposure for all-cause mortality and lung cancer mortality.

We estimated ϕ using g-estimation. G-estimation is an iterative search for the value of ϕ at which T^0 is independent of monthly radon exposure X_k, conditional on covariates. Testing the conditional independence of T^0 and X_k can be done by including the potential outcome, T^0, as an individual level covariate in a model that predicts monthly exposures (the "exposure model"), conditional on prior covariates. The coefficient for T^0 in the exposure model can be used to test this conditional independence. At the estimate of ϕ, monthly radon exposure within groups of similar individuals should not be associated with T^0. A point estimate and associated 95% confidence intervals for ϕ was obtained using a grid-search over a range of values for ϕ (20). Under our model, a TR greater than 1 indicates a harmful exposure.

We model exposure using a log-linear model with modifications to account for unexposed individuals. Our exposure model includes terms for employment status, previous radon exposure during follow-up, race, year of birth, radon exposure before follow-up, years of employment before follow-up, and year of hire. Covariate coding for our exposure model is given in Web Appendix 1 and further technical details regarding our approach to estimating the TR are shown in Web Appendix 2.

Assessing the presence of healthy worker survivor bias:

SNAFT models can adjust for time-varying confounding due to current employment status and history of prior employment status and exposure, which we hypothesized would control healthy worker survivor bias. Following previous authors, current employment status was controlled for by restricting the exposure model to periods of active employment (i.e. $L_k = 1$) (22, 23) and we also adjust for exposure and employment history $(\bar{X}_k, \bar{L}_k)$ by including terms for the history variables described in Web Appendix 1 up to, and including time k. The exposure model may be restricted to specific time periods (such as employed person time) without placing the same restriction on the SNAFT model (24). We refer to this model as our "adjusted" SNAFT model.

We also fit an "unadjusted" SNAFT model that does not adjust for time-varying confounding. The exposure model for the "unadjusted" SNAFT model was used to estimate the expected cumulative exposure (rather than monthly exposures), conditional only on age and the covariates fixed at the beginning of follow-up.

To quantify the magnitude of the healthy worker survivor bias in all models, we report the percent difference between "adjusted" and "unadjusted" models calculated as $100\% * (\phi_{adjusted} - \phi_{unadjusted})/\phi_{adjusted}$. A negative value was interpreted as evidence that the radon-mortality association is underestimated due to healthy worker survivor bias.

We also describe variation in radon-lung cancer association with time since exposure, similar to previous analyses. Using the model shown in (2) we estimated the TR for windows of exposure from the preferred model of the Committee on the Biological Effects of Ionization Radiation (5).

$$T = m + \int_m^{T^0} (1 + \phi_1 \bar{X}_{k1} + \phi_2 \bar{X}_{k2} + \phi_3 \bar{X}_{k3})^{-1} \, dk \qquad (2)$$

In model (2), we let $\bar{X}_{k1}$, $\bar{X}_{k2}$, and $\bar{X}_{k3}$ correspond to the exposure accrued (since follow-up began) between 5-14 years, 15-24, and 25+ years prior. This approach utilizes the same exposure model as we used in our primary analysis. Note that model (1) is a special case of model (2) when $\phi_1 = \phi_2 = \phi_3$.

Our analytic dataset includes both prevalent (miners already employed at study entry) and incident hires (miners who were enrolled in the study at the time they started mining). Because prevalent and incident hires may differ with respect to health status at time of entry into follow-up (25), we assessed the impact of including long-term prevalent hires by restricting models to miners that worked <20, <10, <5, <2.5 or 0 years before enrollment. In these models we collapsed birth cohort from eight to four time periods: <1910 (ref), 1910-1919, 1920-1929, >1929.

SNAFT models are one valid approach for cohort analyses of cumulative exposure-mortality associations under certain conditions, namely when prior exposure affects employment status, and employment affects subsequent exposure and disease. Following previous reports (23), we assessed whether these conditions hold by fitting two standard proportional hazards models. First, we estimated whether prior exposure affects current employment status by fitting a model adjusted for baseline covariates and employment history. Second, we fit a model to compare the hazard of death between person-time not employed and person-time employed (referent) as a uranium miner, adjusted for covariates including cumulative exposure with a lag of 2 years.

RESULTS

Demographics and exposure distribution

Our cohort comprised 4124 white and non-white miners with over 130,000 person-years of follow-up (Table 1). No cause of death could be determined for 22 miners and 14 were lost to follow-up before 1979. A majority of the miners died before 12/31/2005, and a higher proportion of whites than non-whites died of lung cancer (a difference previously attributed to differences in smoking patterns) (26). Non-white miners were followed-up for longer and worked longer during follow-up than white miners, despite similar employment time before follow-up. Across both racial groups, employment duration (as well as radon exposure duration), median monthly exposure (in WLM) among employed person-months, and cumulative exposure (in 100 WLM) at baseline and over follow-up were higher in those who eventually developed lung cancer than in non-cases. Median cumulative exposure was higher during follow-up than prior to first interview. Monthly exposure distributions were highly right skewed and varied with calendar period (Figure 2).

Dose-response analyses

Using a model for all-cause mortality under a 5-year cumulative radon exposure lag, the adjusted TR was higher than the unadjusted TR by 74% (Table 2). Based on our adjusted model, we estimate that, among 3,120 miners who died during follow-up, occupational radon exposure after enrollment was associated with 10,118 person-years of life lost due to premature death (not shown).

For lung cancer, the adjusted TR was higher than the unadjusted TR by 39% (Table 2). Based on our adjusted model, we estimate that, among 617 lung cancer cases, exposure accounted for 6,071 person-years of life lost (not shown). The adjusted TR (95% confidence interval) for lung cancer decreased with time since exposure (Table 3).

After excluding people who had long durations of employment prior to entering the cohort, the adjusted TR decreased relative to the TR in the full cohort (Table 4).

The hazard for terminating employment was lower in workers with cumulative radon exposure above the median (1.2 X 100 WLM) vs. those with cumulative exposure less than the median (referent); HR (95% confidence interval) = 0.90 (0.84, 0.98), not shown. The direction of this association agrees with previous analyses of occupational cohorts using similar or identical statistical models (9, 23, 27). The adjusted hazard of death was higher among person-time not employed relative to employed person-time as a uranium miner; HR

(95% confidence interval) = 3.3 (2.4, 4.3). Thus, regression models adjusting for employment history would be biased and SNAFT models are needed to appropriately adjust for time-varying confounding by employment status.

DISCUSSION

Healthy-worker survivor bias can occur in occupational studies when exposure accrues over time and workers with stronger health and therefore better cancer prognoses remain employed longer. The estimates of the TR were lower in unadjusted models relative to the models adjusted for healthy-worker survivor bias for lung cancer (39%) and all cause mortality (74%). These findings support previous speculation of substantial survivor bias in the Colorado Plateau uranium miner data (28). We observed that prior radon exposure is associated with leaving employment, which multivariable regression models cannot address. SNAFT models can adequately control healthy worker survivor bias in this scenario because the models achieve confounder control without stratification (13). We show that this bias leads to underestimating the slope of the dose-response between radon and both lung cancer and all-cause mortality, which underlie projections of population excess-mortality due to radon exposure.

Previous analyses of miner data may be subject to uncontrolled or improperly controlled healthy worker survivor bias. For example, in their most recent report, the Committee on the Biological Effects of Ionizing Radiation based risk estimates on the so-called “exposure-age-duration” and “exposure-age-concentration” Poisson regression models. These models estimate the relative rate per 100 WLM of radon exposure, stratified on age-at-exposure, attained age, and duration (or concentration) of exposure. Exposure duration is a strong proxy for employment history. Under our hypothesis, risk parameters from the “exposure-age-concentration” and those from the “exposure-age-duration” model may be biased downward. Our findings suggest a stronger healthy worker survivor bias among all causes, perhaps because of the inclusion of causes of death in which healthy worker survivor bias is stronger.

Our estimate of the TR and hazard/rate ratios from previous analyses in this cohort are not directly comparable because mortality rates are not constant over time. Accordingly, we compared adjusted and unadjusted models to assess the magnitude of bias. Other authors have assessed this bias by transforming the TR from SNAFT models to a hazard ratio to compare with results from proportional hazards regression models (22) or parametric accelerated failure time models (29). Our novel approach allows a straightforward comparison of two SNAFT models. However, our approach may be more sensitive to misspecification of exposure models, which are needed for g-estimation. Previous examples have used simpler exposure models than our own, by fitting models for binary exposures (20, 22, 30, 31, 32, 33, 34, 35) or exposure quantiles (29). In contrast, we report SNAFT models under a parametric model for the unbinned exposure (36). In Web Appendix 4, we also fit SNAFT models under alternative exposure models and note that results are somewhat sensitive to the choice of model. We also compared a log-linear SNAFT model to a baseline adjusted parametric accelerated failure time model, which yielded a similar magnitude for healthy worker survivor bias as our approach (Web Appendix 5). SNAFT results had narrower confidence intervals than the parametric model, reflecting different parametric assumptions made by the two approaches.

A second innovation is our use of SNAFT models to explore variation in the time ratio by exposure windows. Such models have been previously proposed in principal (e.g. model 23.10 in Robins and Hernán (2009) (37)), but have not been used in analysis. The TR for each window of exposure can be interpreted as a direct effect of exposure within that time period, not mediated by subsequent exposure (38).

Our analysis is concerned mainly with reducing healthy worker survivor bias, which we conceptualize as a specific instance of time-varying confounding. We also address other sources of variation in the TR in occupational studies, such as left truncation (39). Within our data, miners hired before the study inception in 1950 may be systematically different from the miners who were hired after the study began. As one way to address these possible differences, we adjusted for pre-enrollment exposure and employment history as time-fixed covariates and used them only for control of confounding. Additionally, we did not consider individuals at-risk during the pre-enrollment person-time, which should be considered immortal person time (40). To illustrate the potential bias, we repeated our SNAFT analysis with lung cancer but included immortal person time and pre-enrollment exposures in the cumulative exposure metric. This change resulted in a 34% decrease in the value of ϕ for the adjusted model (not shown).

Another way to address concerns about including data from before study enrollment is to consider differences between “prevalent” and “incident” hires (39). As shown in Table 4, the apparent magnitude of the radon-lung cancer exposure-response decreases after excluding workers with long periods of employment before follow-up. This result runs counter to expectation under the assumption that susceptible individuals will be under-represented in the full cohort. The result may reflect exposure measurement quality changes over time or modification by exposure concentration. We also observed stronger apparent healthy worker survivor bias among prevalent hires (not shown). Both observations may be partly explained by the longer duration of employment during follow-up by prevalent hires (median 4.5 years, not shown) than incident hires (median 3.8 years). Incident hires comprised only 10% of the workforce (n=389, 34 lung cancer deaths; not shown), so inference regarding biases in this group is subject to greater uncertainty.

Confidence intervals are narrower in analyses excluding miners with 5 or more years of employment before enrollment, versus analyses with fewer excluded miners (Table 3). This observation may be due to the reduction in variation of other risk factors for lung cancer that vary by year of hire, such as smoking. In the miner data, we observed that never-smokers were more prevalent among miners hired after 1955 (27%) versus before 1940 (14%) or those hired between 1940 and 1955 (22%, not shown). We did not have access to dates of starting or cessation of smoking (17) and could not evaluate the role of smoking as a time-varying confounder. Previous analyses have suggested that smoking may modify the radon-lung cancer association (41), but is not a source of strong time-fixed (17, 42) or time-varying confounding (43). In our context, smoking may affect both employment status and the outcomes under study (44). SNAFT models can adequately control this bias by adjustment for employment history, if we assume that smoking is not associated with exposure, independent of employment history and the baseline covariates. This assumption may be violated if individuals who start smoking are preferentially placed in lower (or higher) exposed jobs within the mine. This phenomenon would likely present as apparent time-fixed confounding by smoking, as well, which suggests that any residual confounding by smoking is small.

We have mainly addressed issues of confounding by time-varying factors in this analysis. However, the effects of cumulative exposure to radon may be heterogeneous over other time-varying covariates, such as exposure concentration or time since exposure (43, 45). As we have shown, SNAFT models are well suited to address questions regarding time-varying covariates. Unfortunately, our algorithm for a SNAFT model to quantify modification of the TR by exposure concentration did not converge, so we were unable to assess the TR over levels of exposure concentration (not shown). Recent analyses suggest that apparent modification by exposure concentration may be partially due to changes in exposure measurement quality over time (46), which we address in Web Appendix 4. Allowing for modification of the TR would be essential for comparing hypothetical interventions (37) such as more stringent occupational exposure limits (47). This problem echoes previous difficulties with addressing modification in SNAFT models raised by Joffe et al, and may be a shortcoming of using SNAFT models in practice (48). However, our models using time windows of exposure agreed qualitatively with previous analyses (49, 50) suggesting that SNAFT models may be useful for estimating more complex dose-time-response relationships in epidemiologic data.

Conclusion

While we address one kind of bias, any study using miner data is subject to other biases from exposure measurement error that reduces our ability to control confounding (51), and biases the dose response (52), co-exposure to other lung carcinogens such as arsenic (53), diesel exhaust, or silica (54) and reliance on death certificate data. The relative impact of these issues for SNAFT models (compared to regression) is unknown. Further refinement of analyses to include possible dose-response modification by exposure concentration, possibly using pooled data, may better inform risk projection. We show evidence of healthy worker survivor bias in a cohort that plays a key role in risk projection models, and improved handling of employment history as a confounder is a necessary step in reducing this bias.

ACKNOWLEDGEMENTS

Author affiliations: Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina (Alexander P Keil, David B Richardson, and Melissa A. Troester)

Contributions: All listed authors contributed substantially and equally to this work.

This work was supported by the National Institute of Occupational Health Sciences (Grant # T42OH008673-08).

We thank Dr. Stephen R. Cole, Dr. Steve Wing and Dr. Michael Hudgens for expert advice.

Conflicts of interest: none declared.

References

1. Krewski D, Lubin JH, Zielinski JM, Alavanja M, Catalan VS, Field RW, Klotz JB, Létourneau EG, Lynch CF, Lyon JI, and others. Residential radon and risk of lung cancer: a combined analysis of 7 North American case-control studies. *Epidemiology*. 2005; 16(2):137-145.

2. Darby S, Hill D, Auvinen A, Barros-Dios JM, Baysson H, Bochicchio F, Deo H, Falk R, Forastiere F, Hakama M, Heid I, Kreienbrock L, Kreuzer M, Lagarde F, Mäkeläinen I, Muirhead C, Oberaigner W, Pershagen G, Ruano-Ravina A, Ruosteenoja E, Rosario AS, Tirmarche M, Tomásek L, Whitley E, Wichmann H, and Doll R. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. *Brit Med J*. 2005; 330(7485):223-226.

3. United States Environmental Protection Agency. *EPA Assessment of Risks from Radon in Homes*. Washington, DC: Office of Radiation and Indoor Air; 2003. (EPA no 402-R-03-003).

4. National Research Council. *Report of the committee on the biological effects of ionizing radiation: Health Risks Of Radon And Other Internally Deposited Alpha-Emitters: BEIR IV*. Committee on Health Risks of Exposure to Radon (BEIR IV) . Board on Radiation Effects Research, Commission on Life Sciences. Washington, DC: National Academy Press 1988.

5. National Research Council. *Health effects of exposure to radon, BEIR VI*. Committee on Health Risks of Exposure to Radon (BEIR VI). Board on Radiation Effects Research, Commission on Life Sciences. Washington, DC: National Academy Press, 1999.

6. World Health Organization. *WHO Handbook on Indoor Radon. A Public Health Perspective*. Geneva, Switzerland: World Health Organization; 2009.

7. ICRP. ICRP Publication 115: Lung cancer risk from radon and progeny and statement on radon. *Ann ICRP*. 2010; 40(1):1-64.

8. Arrighi HM and Hertz-Picciotto I. The evolving concept of the healthy worker survivor effect. *Epidemiology*. 1994; 5(2):189-196.

9. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period - application to control of the healthy worker survivor effect. *Math Mod*. 1986; 7(9):1393-1512.

10. Hornung RW, Deddens JA, and Roscoe RJ. Modifiers of lung cancer risk in uranium miners from the Colorado Plateau. *Health physics*. 1998; 74(1):12-21.

11. Hornung RW, Deddens J, and Roscoe R. Modifiers of exposure-response estimates for lung cancer among miners exposed to radon progeny. *Environ Health Perspect*. 1995; 103(Suppl 2):49-53.

12. Pearce N. Time-related confounders and intermediate variables. *Epidemiology*. 1992; 3(4):279-281.

13. Robins JM. The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: Sechrest L, Freeman H, Mulley A, editors. *Health Service Research Methodology: A Focus on AIDS*. Washington, DC: National Center for Health Services Research, US Public Health Service; 1989. p. 113-159.

14. Robins JM and Tsiatis AA. Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika*. 1992; 79(2):311-319.

15. Holaday DA, Rushing DE, Coleman RD, Woolrich PF, and Kusnetz HL. Control of radon and daughters in uranium mines and calculations on biologic effects. Washington, DC: US Department of Health, Education and Welfare, Public Health Service; 1957. (Public Health Service Publication No. 494).

16. Roscoe R. An update of mortality from all causes among white uranium miners from the Colorado Plateau study group. *American Journal of Industrial Medicine*. 1997; 31(2):211-222.

17. Schubauer-Berigan MK, Daniels RD, and Pinkerton LE. Radon exposure and mortality among white and American Indian uranium miners: an update of the Colorado Plateau cohort. *Am J Epidemiol*. 2009; 169(6):718-730.

18. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*.New York, NY: Wiley; 1980.

19. Lubin JH. Models for the analysis of radon-exposed populations. *Yale J Biol Med*. 1988; 61(3):195-214.

20. Hernán MA, Cole SR, Margolick J, Cohen M, and Robins JM. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidem Dr S*. 2005; 14(7):477-491.

21. Picciotto S, Chevrier J, Balmes J, and Eisen EA. 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.

22. Chevrier J, Picciotto S, and Eisen EA. Chevrier J, Picciotto S, and Eisen EA. A comparison of standard methods with g-estimation of accelerated failure-time models to address the healthy-worker survivor effect: application in a cohort of autoworkers exposed to metalworking fluids. *Epidemiology*. 2012; 23(2):212-219.

23. Naimi AI, Cole SR, Hudgens MG, Brookhart MA, and Richardson DB. Assessing the component associations of the healthy worker survivor bias: occupational asbestos exposure and lung cancer mortality. *Ann Epidemiol*. 2013; 23(6):334-341.

24. Joffe MM, Yang WP, and Feldman HI. Selective ignorability assumptions in causal inference. *Int J Biostat*. 2010; 6(2):1-25.

25. Applebaum KM, Malloy EJ, and Eisen EA. Left truncation, susceptibility, and bias in occupational cohort studies. *Epidemiology*. 2011; 22(4):599-606.

26. Roscoe RJ, Deddens JA, Salvan A, and Schnorr TM. Mortality among Navajo uranium miners. *Am J Public Health*. 1995; 85(4):535-540.

27. Cole SR, Richardson DB, Chu , and Naimi AI. Analysis of Occupational Asbestos Exposure and Lung Cancer Mortality Using the G Formula. *Am J Epidemiol*. 2013; 177(9):989-996.

28. Stayner L, Steenland K, Dosemeci M, and Hertz-Picciotto I. Attenuation of exposure-response curves in occupational cohort studies at high exposure levels. *Scand J Work Environ Health*. 2003; 29(4):317-324.

29. Naimi AI, Cole SR, Hudgens MG, and Richardson DB. Estimating the effect of cumulative occupational asbestos exposure on time to lung cancer mortality: using structural nested failure-time models to account for healthy-worker survivor bias. *Epidemiology*. 2014; 25(2):246-254.

30. Robins JM, Blevins D, Ritter G, and Wulfsohn M. G-estimation of the effect of prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology*. 1992; 3(4):319-336.

31. Witteman JC, D'Agostino RB, Stijnen T, Kannel WB, Cobb JC, de Ridder MA, Hofman A, and Robins JM. G-estimation of causal effects: isolated systolic hypertension and cardiovascular death in the Framingham Heart Study. *Am J Epidemiol*. 1998; 148(4):390-401.

32. Joffe MM, Hoover DR, Jacobson LP, Kingsley L, Chmiel JS, Visscher BR, and Robins JM. Estimating the effect of zidovudine on Kaposi's sarcoma from observational data using a rank preserving structural failure-time model. *Stat Med*. 1998; 17(10):1073-1102.

33. Keiding N, Filiberti M, Esbjerg S, Robins JM, and Jacobsen N. The graft versus leukemia effect after bone marrow transplantation: A case study using structural nested failure time models. *Biometrics*. 1999; 55(1):23-28.

34. Korhonen PA, Laird NM, and Palmgren J. Correcting for non-compliance in randomized trials: an application to the ATBC Study. *Stat Med*. 1999; 18(21):2879-2897.

35. Cole SR and Chu H. Effect of acyclovir on herpetic ocular recurrence using a structural nested model. *Contemp Clin Trials*. 2005; 26(3):300-310.

36. Li N, Elashoff DA, Robbins WA, and Xun L. A hierarchical zero-inflated log-normal model for skewed responses. *Stat Methods Med Res*. 2011; 20(3):175-189.

37. Robins JM, Hernan MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G, eds. *Longitudinal Data Analysis*. Boca Raton, FL: CRC Press; 2009: 553–599.

38. Westreich D and Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *Am J Epidemiol*. 2013; 177(4):292-298.

39. Applebaum KM, Malloy EJ, and Eisen EA. Reducing healthy worker survivor bias by restricting date of hire in a cohort study of Vermont granite workers. *J Occup Environ Med*. 2007; 64(10):681-687.

40. Suissa S. Immortal time bias in pharmacoepidemiology. *Am J Epidemiol*. 2008; 167(4):492-499.

41. Tomásek L. Interaction of radon and smoking among Czech uranium miners. *Radiat Prot Dosimetry*. 2011; 145(2-3):238-242.

42. Leuraud K, Schnelzer M, Tomásek L, Hunter N, Tirmarche M, Grosche B, Kreuzer M, and Laurier D. Radon, smoking and lung cancer risk: results of a joint analysis of three European case-control studies among uranium miners. *Radiat Res*. 2011; 176(3):375-387.

43. Richardson DB, Cole SR, and Langholz B. Regression Models for the Effects of Exposure Rate and Cumulative Exposure. *Epidemiology*. 2012; 23(6):892-899.

44. Pearce N, Checkoway H, and Kriebel D. Bias in occupational epidemiology studies. *Occup Environ Med*. 2007; 64(8):562-568.

45. Lubin JH, Boice JD Jr, Edling C, Hornung RW, Howe G, Kunz E, Kusiak RA, Morrison HI, Radford EP, and Samet JM. Radon-exposed underground miners and inverse dose-rate (protraction enhancement) effects. *Health Phys*. 1995; 69(4):494-500.

46. Allodji RS, Leuraud K, Thiébaut ACM, Henry S, Laurier D, and Bénichou J. Impact of measurement error in radon exposure on the estimated excess relative risk of lung cancer death in a simulated study based on the French Uranium Miners' Cohort. *Radiat Environ Biophys*. 2012; 51(2):151-163.

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

48. Joffe MM, Yang WP, and Feldman H. G-estimation and artificial censoring: problems, challenges, and applications. *Biometrics*. 2012; 68(1):275-286.

49. Langholz B, Thomas D, Xiang A, and Stram D. Latency analysis in epidemiologic studies of occupational exposures: application to the Colorado Plateau uranium miners cohort. *Am J Ind Med*. 1999; 35(3):246-256.

50. Tomásek L. Lung cancer mortality among Czech uranium miners-60 years since exposure. *J Radiol Prot*. 2012; 32(3):301-314.

51. Armstrong BG. Effect of measurement error on epidemiological studies of environmental and occupational exposures. *J Occup Environ Med*. 1998; 55(10):651-656.

52. Stram DO, Langholz B, Huberman M, and Thomas DC. Correcting for exposure measurement error in a reanalysis of lung cancer mortality for the Colorado Plateau Uranium Miners cohort. *Health Phys*. 1999; 77(3):265-275.

53. Arrighi HM and Hertz-Picciotto I. Controlling the healthy worker survivor effect: an example of arsenic exposure and respiratory cancer. *Occup Environ Med*. 1996; 53(7):455-462.

54. Bergdahl IA, Jonsson H, Eriksson K, Damber L, and Järvholm B. Lung cancer and exposure to quartz and diesel exhaust in Swedish iron ore miners with concurrent exposure to radon. *J Occup Environ Med*. 2010; 67(8):513-518.

Table 1. Demographics, Follow-up Characteristics and Radon Exposures, 4134 Male Uranium Miners, Colorado Plateau, USA 1950-2005

Characteristic	Race					
	White[a] (N=3,335)			Other[a] (N=769)		
	No.	%	Median (IQR)	No.	%	Median (IQR)
Vital status[b]						
Alive	790	23.5		214	27.8	
Deceased (cause unknown)	51	1.5		20	2.6	
Deceased (known cause)	2514	74.9		535	69.6	
Deceased (lung cancer)	554	16.5		63	8.1	
Birth cohort						
<1900	171	5.1		21	2.7	
1900-1909	460	13.7		76	9.8	
1910-1919	857	25.2		131	16.8	
1920-1929	890	26.5		284	36.5	
1930-1939	890	26.5		258	33.1	
1940-1949	87	2.6		9	1.2	
Date of hire, year						
Cases[c]			1953 (1950, 1956)			1953 (1951, 1956)
Non-cases			1955 (1952, 1957)			1954 (1951, 1957)
Total			1954 (1951, 1957)			1954 (1951, 1957)
Years of follow-up						
Cases[b]			28.0 (18.7, 37.4)			31.1 (24.7, 40.3)
Non-cases			35.9 (19.8, 45.6)			39.8 (26.3, 48.5)
Total			34.1 (19.5, 45.5)			38.6 (26.2, 47.1)
Active employment years during follow-up						
Cases[c]			7.4 (3.6, 10.9)			10.8 (7.5, 12.5)
Non-cases			3.5 (0.8, 7.7)			5.6 (1.5, 8.9)
Total			4.0 (1.0, 8.2)			5.6 (1.5, 9.6)
Active employment years at entry[d]						
Cases[c]			2.4 (0.79, 6.0)			2.5 (1.1, 4.0)
Non-cases			1.3 (0.30, 3.9)			1.2 (0.21, 3.0)
Total			1.5 (0.29, 4.0)			1.4 (0.29, 3.0)
Monthly exposure among active work time (WLM)						
Cases[c]			4.5 (2.4, 9.0)			3.8 (2.5, 8.1)
Non-cases			3.1 (1.4, 6.6)			2.6 (1.1, 5.8)
Total			3.4 (1.6, 7.2)			2.9 (1.2, 6.1)
Cumulative Radon during follow-up (100 WLM)						
Cases[c]			4.6 (1.8, 9.5)			6.2 (3.3, 11.1)
Non-cases			1.6 (0.44, 4.1)			2.0 (0.6, 5.2)
Total U mining			1.9 (0.55, 4.9)			2.4 (0.65, 5.8)
Cumulative Radon at entry[d] (100 WLM)						
Cases[c]			2.7 (0.59, 8.6)			1.7 (0.47, 6.7)
Non-cases			1.0 (0.15, 3.9)			0.68 (0.11, 2.3)
Total U mining			1.2 (0.19, 4.6)			0.76 (0.13, 2.7)
Hard rock mining			0.00 (0.00, 0.18)			0.00 (0.00, 0.00)

WLM: working level months; IQR: interquartile range

[a] Total person years is 107,626 for white race and 27,343 for other race

[b] Vital status as of 31 December, 2005

[c] Cases = individuals who died during follow-up with underlying cause of death listed as lung cancer

[d] Entry into follow-up defined as date of first interview by the United States Public Health Service

Table 2. Time Ratio per 100 Working Level Months, Lagged 5 Years, 4134 Male Uranium Miners, Colorado Plateau, USA 1950-2005

	Lung cancer		
Model	**Time ratio**	**95% CI**	**%diff[c]**
Lung cancer			
Adjusted[ab]	1.168	1.152, 1.174	ref
Unadjusted[a]	1.102	1.099, 1.112	-39
All-causes			
Adjusted[ab]	1.054	1.041, 1.068	ref
Unadjusted[a]	1.014	1.013, 1.015	-74

CI: confidence interval; diff: difference

[a] Adjusted for time-fixed covariates: exposure from uranium mining before enrollment, exposure from hard rock mining before enrollment, race, birth cohort, date of hire

[b] Also adjusted for time-varying covariates: annual exposure during follow-up from 1, 2, 3, 4, 5 and cumulative exposure from 6-10 years and 10+ years prior, current employment status, and cumulative time at work during follow-up

[c] Percent difference in ϕ from adjusted model, defined in text

Table 3. Time Ratio per 100 Working Level Months for Windows of Exposure, 4134 Male Uranium Miners, Colorado Plateau, USA 1950-2005

Exposure window[a]	Time Ratio[b]	95% CI
5-14 years	1.188	1.116, 1.230
15-24 years	1.128	1.050, 1.294
25+ years	1.022	0.950, 1.198

CI: confidence interval

[a]Exposure following enrollment accrued within the noted period

[b]Adjusted for exposure from uranium mining before enrollment, exposure from hard rock mining before enrollment, race, birth cohort, date of hire, annual exposure during follow-up from 1, 2, 3, 4, 5 and cumulative exposure from 6-10 years and 10+ years prior, current employment status, and cumulative time at work during follow-up

Table 4. Sensitivity Analysis for Inclusion of Prevalent Hires in the Study Cohort on the adjusted Time Ratio for Radon/Lung Cancer Association, 4134 Male Uranium Miners, Colorado Plateau, USA 1950-2005

Max. employment prior to enrollment[a]	Time ratio[b]	95% CI
Full cohort[c]	1.095	1.087, 1.117
20 years	1.092	1.087, 1.112
10 years	1.094	1.085, 1.114
5 years	1.086	1.075, 1.089
2.5 years	1.082	1.074, 1.088
Incident hires only	1.070	1.063, 1.076

CI: confidence interval

[a] For each row, workers were excluded if they worked longer than this amount before study enrollment,

[b]Per 100 working level months; adjusted for: baseline exposure from uranium mining, race, prior mining exposure, birth cohort, date of hire, annual exposure during follow-up from 1, 2, 3, 4, 5 and cumulative exposure from 6-10 years and 10+ years prior active employment status, and cumulative time at work during follow-up.

[c] Birth cohort was represented by 4 groups, resulting in different time ratios between analysis with no exclusions and results from table 3 (in which birth cohort was represented by 8 groups).

Figure Legends

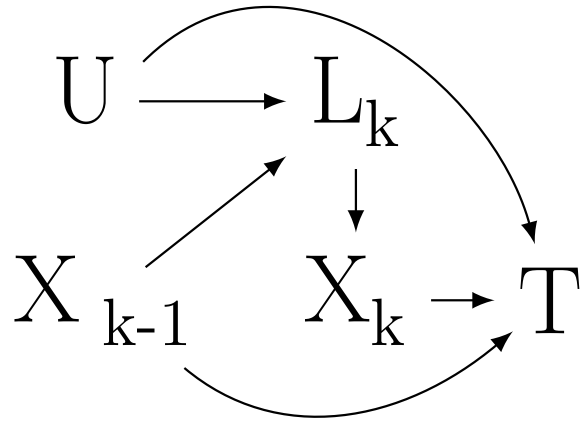

Figure 1: Causal diagram showing hypothesized relationships underlying control healthy worker survivor bias in the Colorado Plateau Uranium Miners data. Confounding of the association between radon exposure, X_k and age at death T occurs through employment status L_k in month k, possibly by an unmeasured predictor of leaving employment and death, U. Stratifying on L_k in a regression model induces bias in the coefficient for prior radon exposure X_{k-1}.

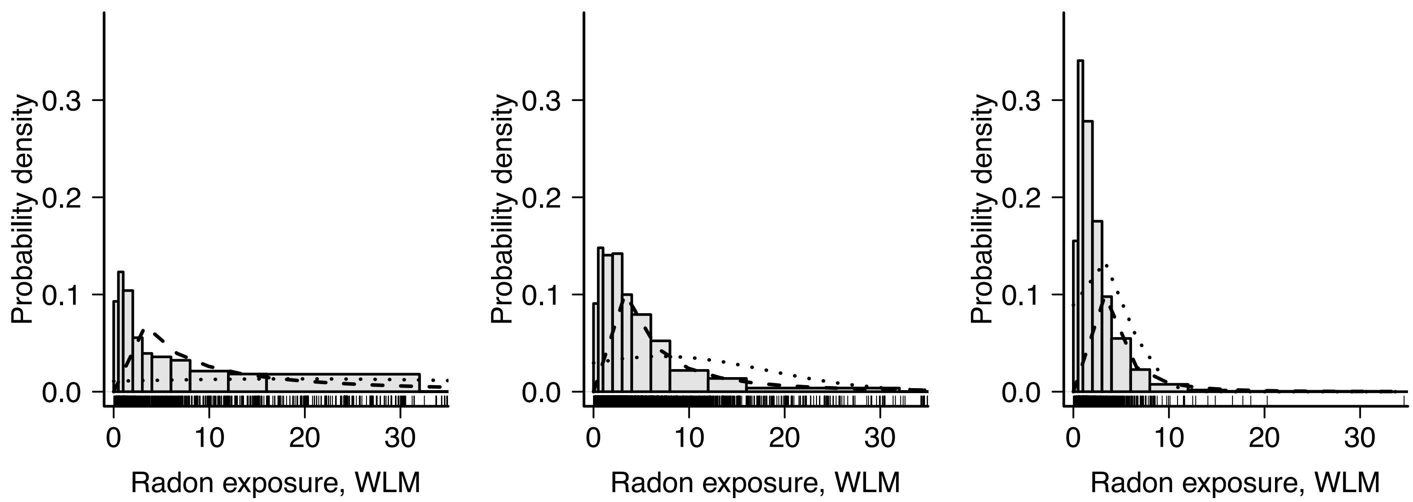

Figure 2: Illustrative monthly exposure distributions for white males born from 1920-1929. Months selected represent the 95th (panel A, 95th percentile), 50th (B, median), and 5th (C, 5th percentile) percentiles of the mean monthly exposure from 1950-1969. Figures show histograms with cut-points at 0, 1, 2, 3, 4, 6, 8, 12, 16, 32, 32+ and normal (dashed lines) and log-normal (dotted lines) curves fit to data. Lines below histogram represent monthly exposures for individual miners. Exposures truncated at 35 WLM (working level months). Colorado Plateau, USA, 1950-2005.