HEALTHY WORKER SURVIVOR BIAS IN A COHORT OF URANIUM MINERS FROM THE COLORADO PLATEAU

Alexander P Keil

A dissertation submitted to the faculty of the University of
North Carolina at Chapel Hill in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department
of Epidemiology

Chapel Hill
2014

Approved by:

David Richardson

Stephen Cole

Michael Hudgens

Melissa Troester

Steven Wing

ABSTRACT

ALEXANDER KEIL: HEALTHY WORKER SURVIVOR BIAS IN A COHORT OF URANIUM MINERS FROM THE COLORADO PLATEAU
(Under the direction of David Richardson)

Radon, a ubiquitous gas present in breathing air and concentrated in the indoor environment, is a well established risk factor for lung cancer. Primarily, evidence for this association originated in studies of miners occupationally exposed to high concentrations of radon. Much work has been done to predict lung cancer risk due to lower dose exposures in residences using dose-response curves derived from long-term, high-dose miner studies and shorter-term, low-dose residential studies. While residential studies suffer from a high probability of exposure misclassification at low exposures, miner studies present an opportunity to apply more precise estimates of the lung cancer-radon association to risk assessments. However, potential bias due to the Healthy Worker survivor bias has not been addressed in previous studies of occupational exposure to radon. The Healthy Worker survivor bias occurs when workers with poor prognosis leave work sooner than those with better prognosis, thus creating an apparent association between low cumulative occupational exposures and mortality. Healthy worker survivor bias has been shown to substantially bias dose-response estimates in other settings, but it has not been explored in occupational studies of radon exposure. We apply two g-methods designed for addressing healthy worker survivor bias that cannot be controlled using conventional statistical methods. We utilize data from the Colorado Plateau uranium miners cohort, which comprises 4,137 male uranium miners who agreed to participate in a health study between 1950 and 1960 and were followed up for mortality through 2005. Our results suggest that there may be

healthy worker survivor bias of the association between cumulative radon exposure and both lung cancer and all cause mortality. This work highlights the need for non-standard approaches to controlling time-varying confounding in occupational data. We show that, under certain conditions, g-methods can control this confounding, but that careful consideration should be made in the choice of method.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF ABBREVIATIONS

AFT Accelerated failure time

BEIR Biological effects of ionizing radiation

Bq Becquerel

CEA Commissariat á l'énergie

COGEMA Compagnie Générale des Matiéres Nucléaires

CPUM Colorado Plateau uranium miners

ERR Excess relative risk/rate

HR Hazard ratio

HWSB Healthy worker survivor bias

IARC The International Agency for Research on Cancer

ICRP International Commission on Radiological Protection

IP, IPW inverse probability, inverse probability weights

LET Linear energy transfer

LNT Linear-no threshold

MSM Marginal structural model

NCRP National Council on Radiological Protection

RR Relative risk/rate

SI The International System of Units (Le Système International d'unités)

SNAFT Structural nested accelerated failure time

TSE Time since exposure

UNSCEAR United Nations Scientific Committee on the Effects of Atomic Radiation

USEPA United States Environmental Protection Agency

WL Working level

WLM Working level month

CHAPTER I: INTRODUCTION

1.1 Radon

1.1.1 Public health significance

Elemental radon is a naturally occurring gas at room temperature and is carcinogenic to humans when it is part of the breathing air due to the radioactivity of both radon and its progeny (*IARC* (2012)). The most common isotope, ^{222}Rn and its progeny shown in Figure 2 1.1 contribute most of the radiation dose (*IARC* (1988, 2001); *ICRP* (2010); *Chen et al.* (2014)). Because radon is ubiquitous in air, soil, and water and concentrates in indoor air, there is substantial exposure potential in populations who spend time indoors *NRC* (1999). This problem may be compounded by the use of certain building materials, which also release radon (*Gierl et al.* (2014); *Zhukovsky and Vasilyev* (2014)).

^{222}Rn has an unstable nucleus and decays with a half-life of around four days by releasing an α particle (a particle with two protons and two neutrons) and starting a chain of rapid, α and β particle releasing decays from what are known as the radon progeny (historically, radon daughters, figure 1.1), which can be inhaled directly or can adhere to airborne particles that are also inhaled. The α particles are sources of high linear-energy-transfer (high-LET) ionizing radiation *ICRU* (2011c,a), resulting in more concentrated cellular damage than some other types of radiation, such as γ radiation (e.g. from atomic bombs dropped on Japan in World War II *Preston et al.* (2003)) or β radiation from radio-iodine (e.g. from releases following the nuclear power accident in Chernobyl (*Brenner et al.* (2011))), provided that the α particles are exposed to sensitive

tissue, such as lung tissue.

The Committee to Study the Biological Effects of Ionizing Radiation (BEIR VI committee) estimates that radon in indoor air is the second leading cause of lung cancer and plays a role in 15,000 to 22,000 lung cancer deaths per year in the United States, though risk projections are consistent with estimates from 3,000 to 33,000 (*NRC* (1999)). Unfortunately, increasing energy efficiency in homes may be resulting in an increase in indoor radon concentrations (*Jiránek and Kačmařikova* (2014); *Yarmoshenko et al.* (2014)), which may exacerbate this problem. Importantly, radon exposures can be reduced through measures such as home radon mitigation (*Steck* (2012)). Thus, the public health impact of radon on lung cancer is high, we may be heading towards an increase in the average population exposure, but there are possible interventions that can help reduce the burden of health impacts from radon exposure. However, because radon mainly comes from the soil, we can realistically only hope to reduce, rather than eliminate exposure, so epidemiologists have a role in estimating the public health impact of radon.

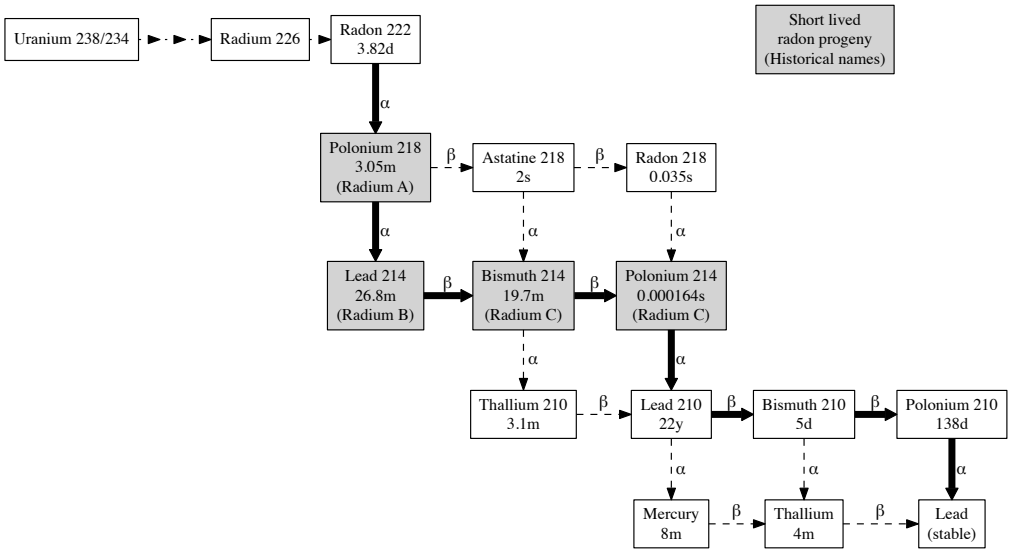

Figure 1.1: ^{222}Rn Decay Chain from ^{238}U or ^{234}U. Thoron (^{220}Rn) and acton (^{219}Rn) produce different progeny.

As noted above, there are many lung cancer deaths in the US that are possibly attributable to radon. However, radon exposure occurs over the entire life course, and biologically meaningful exposure can potentially occur over that whole period. Thus, the health effects of radon are difficult to study without having some measurement or estimate of radon exposure over a long period of time. Factors such as the age at exposure, the intensity or rate of exposure, the duration of exposure, and smoking may all modify the degree to which radon exposure increases the rate of lung cancer (*NRC* (1999)), which only increases this difficulty. To simplify the analysis, radon exposure has often been operationalized using a time-integrated metric of exposure, such as cumulative exposure. Cumulative exposure combines exposure rate with exposure duration to form a single summary measure across time.

Evidence from animal and human studies suggests that the relative risk of lung cancer from radiation exposure is well summarized by a linear fit with no threshold below which no effects occur (*Brenner* (2009)). Thus, a standard approach in studies of radon exposure is to present a linear dose-response for cumulative exposure. The time-related factors noted above (and discussed in more detail in Appendix C.2) are one reason why cumulative exposure may be too simple to accurately model radon exposure in relation to lung cancer. However, there may be other issues with aggregate exposure metrics that are not strongly appreciated in the radon literature. In the current manuscript, we summarize the some of the lines of research on the public health impact of radon on lung cancer, discuss methodologic issues related to the aggregate exposure metrics used in previous radon research, propose and implement possible solutions to one of those issues, and discuss the implications for our knowledge about the health effects of radon on lung cancer.

1.1.2 Lines of research of the public health impact of radon on lung cancer incidence and mortality

Substantial work has been done to model the population impact of radon exposure and cancers in the respiratory tract and lungs, which are the primary exposure sites for radon *ICRP* (1993). Generally, the lines of research into models of radon carcinogenesis have followed one of three approaches: 1) an empirical or statistical modeling approach, which is focused on explaining macro-level trends in radon induced lung cancer, such as dose-response, dose-rate effects or empirical induction periods, 2) mechanistic modeling, which seek to link observed data directly to biologic processes using biologic models for the observed data (e.g. *Richardson* (2009a)) or 3) the dosimetric approach, which involves extrapolating radiation dose-response curves from other ionizing radiation (particularly those derived from the atomic bomb cohort in Japan, e.g. *Preston et al.* (2007)) (*NRC* (1999)).

Much of the evidence of the human health effects of radon derives from studies of uranium miners, who are exposed to relatively high levels of radon due to the presence of radioactive element deposits in the mines and poor ventilation. The Colorado Plateau uranium miners cohort holds a unique place within this field of study due to its long follow-up and large variation in exposures that allow for precise dose-response measures to be estimated (*NRC* (1999)). Recent monographs by major ionizing radiation committees have advocated the use of residential data with correction for exposure misclassification for summary measures of public health impact and risk assessment (*ICRP* (2010); *UNSCEAR* (2008)), but outstanding questions still exist that can be best answered using miner study data, especially with respect to time-related aspects of the radon-lung cancer association. Previous results based on miners studies have also ignored a potential source bias particular to occupational studies, known as healthy worker survivor bias, which has potentially led to biased risk projections generated by miner data.

1.1.2.1 Miner studies of the radon-lung cancer association

The epidemiologic literature on the radon-lung cancer association in miners has a long history and spans multiple countries with large cohort studies of underground miners, such as those from Canada (e.g. *Lane et al.* (2010)), the Czech Republic (e.g. *Tomásek* (2011)), France (e.g. *Rage et al.* (2012)), Germany (e.g. *Kreuzer et al.* (2012)), Sweden (e.g. *Jonsson et al.* (2010)) and the United States (e.g. *Schubauer-Berigan et al.* (2009)). Several pooled analyses of these (and other) cohorts have been performed in an effort to increase the precision of association parameters (*Lubin* (1994); *NRC* (1988, 1999); *Leuraud et al.* (2011); *Fornalski and Dobrzynski* (2011)). The Committee on the Biological Effects of Ionizing Radiations from the National Research Council (United States) periodically releases summary reports of scientific literature on the health risks due to radiation, including the two reports on the risk of lung cancer due to radon in miners given in *NRC* (1988) and *NRC* (1999). The United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) have summarized miner data in a report to the General Assembly (*UNSCEAR* (2008)). Additionally, the International Commission on Radiological Protection (ICRP) has released two summaries of epidemiologic data pertinent to the radon-lung cancer association *ICRP* (1993, 2010), which are shown in Table 1.1 (adapted from *ICRP* (2010)). Each publication provided a table with the major summary estimates of the excess relative risk per 100 working level months (WLM - defined as any combination of exposure rate in working levels [130,000 MeV of potential α energy per liter of air] and employment time that leads to an average exposure of 1 working level over 170 hours).

As shown in Table 1.1 there is substantial heterogeneity of the estimate of the excess relative rate (ERR, defined as the rate ratio - 1) per 100 WLM within each combined analysis. In the BIER VI report, for example, the ERR per 100 WLM ranged from 0.16 in the Chinese tin miner cohort to 5.1 in the Radium Hill uranium miner cohort. In the BEIR VI report, *NRC* (1999) cited lack of adequate data on lung cancer risk factors as

a reason for heterogeneity in results. The BEIR VI committee was able to estimate the modification of the ERR by age at exposure, duration of exposure, and exposure rate, however, and came up with a set of preferred models, discussed in §1.3 of the current manuscript. In the following, we discuss another possible source of heterogeneity that may derive from differences between cohorts in the complex relationships between employment and health.

1.2 Healthy worker survivor bias

1.2.1 Description

Radon is one of many occupational agents that may have persistent negative effects on health. Exposure from the past may have an effect on current disease status. Exposure duration may also play a role in disease as cellular repair pathways from prior insults may be overburdened and not able to cope with more recent insults. One way of approaching analyses of the health effects of such agents is to aggregate exposure to them over time. Within the miner literature, cumulative exposure to radon, duration of exposure, average exposure, or exposure accrued within specific time windows are all examples of such aggregation. While these are all simplifications of a complex, dynamic process of how exposure may change over time, they are nonetheless very useful for summarizing the effects of exposures and for projecting the effects of exposure in other populations who have long term exposures.

In occupational studies, it is frequently observed that the healthiest workers in a given industry tend to remain employed longer than less healthy workers (*Fox and Collier* (1976)). When this scenario occurs, healthier workers can accrue larger cumulative exposures simply for the fact that they remain employed longer. If the health related factors that influence employment also influence the health outcomes of interest, there may be spurious relationships between aggregated exposures and those

Table 1.1: Summary effect estimates for the radon-lung cancer association in underground miners from 3 major publications

Place	Country	Mine type	Follow Up	No. miners	Avg. Cum. Exposure (WLM)†	Total person-years †§	ERR per 100 WLM	95% CI
Colorado	USA	Uranium	1951–1982	2,975	510	66,237	0.6	0.30–1.42
Ontario	Canada	Uranium	1955–1981	11,076	37	217,810	1.42	0.60–3.33
New Mexico	USA	Uranium	1957–1985	3,469	111	66,500	1.81	0.71–5.46
Beaverlodge	Canada	Uranium	1950–1980	6,895	44	114,170	1.31	0.60–3.01
West Bohemia	Czech Republic	Uranium	1953–1985	4,042	227	97,913	1.7	1.21–2.41
CEA-COGEMA	France	Uranium	1946–1985	1,785	70	44,005	0.6	0.00–1.63
Malmberget	Sweden	Iron	1951–1976	1,292	98	27,397	1.42	0.30–9.57
ICRP - 1993				**31,486**	**120**	**635,022**	**1.34**	**0.82–2.13**
Yunnan	China	Tin	1976–1987	13,649	286	134,842	0.16	0.1-0.2
W-Bohemia	Czech Republic	Uranium	1952–1990	4,320	196.8	102,650	0.34	0.2-0.6
Colorado **	USA	Uranium	1950–1990	3,347	578.6	79,556	0.42	0.3-0.7
Ontario	Canada	Uranium	1955–1986	21,346	31	300,608	0.89	0.5-1.5
Newfoundland	Canada	Fluorspar	1950–1984	1751	388.4	33,795	0.76	0.4-1.3
Malmberget	Sweden	Iron	1951–1991	1,294	80.6	32,452	0.95	0.1-4.1
New Mexico	USA	Uranium	1943–1985	3457	110.9	46,800	1.72	0.6-6.7
Beaverlodge	Canada	Uranium	1950–1980	6,895	21.2	67,080	2.21	0.9-5.6
Port Radium	Canada	Uranium	1950–1980	1,420	243	31,454	0.19	0.1-1.6
Radium Hill	Australia	Uranium	1948–1987	1,457	7.6	24,138	5.06	1.0-12.
CEA-COGEMA	France	Uranium	1948–1986	1,769	59.4	39,172	0.36	0.0-1.2
BEIR VI - 1999				**60,606**	**164.4**	**888,906**	**0.59**	**1.32***
Colorado**	USA	Uranium	1950–1990	3,347	807	75,032	0.42	0.3–0.7
Newfoundland	Canada	Fluorspar	1951–2001	1,742	378	70,894	0.47	0.28–0.65
Yunnan	China	Tin	1976–1987	13,649	277	135,357	0.16	0.1–0.2
Wismut	Germany	Uranium	1946–1998	59,001	242	1,801,626	0.21	0.18–0.24
Malmberget	Sweden	Iron	1951–1990	1,415	81	32,452	0.95	0.1–4.1
West Bohemia	Czech Republic	Uranium	1952–1999	9,979	70	261,428	1.6	1.2–2.2
CEA-COGEMA	France	Uranium	1946–1994	5,098	37	133,521	0.8	0.3–1.4
Ontario	Canada	Uranium	1955–1986	21,346	31	319,701	0.89	0.5–1.5
Beaverlodge	Canada	Uranium	1950–1999	10,050	23	285,964	0.96	0.56–1.56
UNCSCEAR - 2009				**125,627**		**3,115,975**	**0.59**	**0.35–1.00**

Adapted from tables A.1-A.3 in *ICRP* (2010).

WLM, working level month; ERR, excess relative risk; CI, confidence interval

∗Standard errors reported rather than 95%CI

† Includes only person time less than 2000 WLM (ICRP) or 3200 WLM (BEIR VI) cumulative exposure

§ Among exposed (BEIR VI only)

** UNSCEAR and BEIR VI report identical ERR/WLM estimates, but difference in average dose and person years is unexplained

health outcomes. This process is referred to in the occupational epidemiologic literature as healthy worker survivor bias (*Buckley et al.* (2014)).

Healthy worker survivor bias has been traditionally considered simultaneously with a second bias, the healthy hire effect. This second bias describes a cohort selection process whereby unhealthy individuals are less likely to enter into employment (*Arrighi and Hertz-Picciotto* (1994, 1993)). Often, these biases are considered as a part of a single process termed the healthy worker effect. Both of these aspects of the healthy worker effect can be seen as specific instances of unmeasured confounding (*Breslow and Day* (1987)), but their unique characteristics and apparent ubiquity warrant special treatment here.

The healthy hire effect describes an unmeasured difference in health status between the working population and the target population for inference. Because of this, comparisons of disease rates between occupational cohorts and other populations can yield biased estimates of effects of occupational exposures. This issue is circumvented through internal analyses of occupational groups, in which rates of disease are compared across exposure levels within the cohort. If exposure is well characterized, internal analyses can result in an estimate of the dose-response – how the disease rates change over increments of exposure – which can inform occupational limits and public health efforts. However, effect estimates from internal analyses can be subject healthy worker survivor bias.

The healthy worker effect in aggregate has been acknowledged by the BEIR committee as a source of bias in uranium miner studies (*NRC* (1998, 1988)). However, healthy worker survivor bias has not been directly assessed in any major dose-response analysis of the radon-lung cancer association. Several lines of evidence suggest that healthy worker survivor bias may bias dose-response estimation in miner cohorts. Thus, risk projection of miner-based evidence to the general population may be based on biased associations. The work by *Robins* (1986) suggests that the existence of a healthy

worker survivor bias is a structural component to occupational studies and would be expected in any study of the health effects of aggregated exposures within a dynamic workforce. Further, under certain models, healthy worker survivor bias is intractable to conventional statistical methods and may require more advanced approaches (*Arrighi and Hertz-Picciotto* (1994, 1993); *Flanders et al.* (1993); *McNamee* (2003); *Pearce et al.* (2007)).

1.2.1.1 Notation

Throughout the remainder of the current manuscript, we adopt the notation shown in table 1.2.

Table 1.2: Notation used in this document referring to observed quantities

Variable*	Interpretation	Examples
k	A particular point in time	Age 30, 1930
X_k	Exposure of interest at time k**	Radon exposure at age 19
$\bar{\boldsymbol{X}}_{\boldsymbol{k}}$	Exposure history at time k**	Annual exposures up to age 30, Monthly exposures from 1955 to 1960
$\bar{X}_k$	Summary exposure history**	Cumulative radon exposure up to age 30, Mean exposure up to age 30, Exposure duration up to age 30
$\boldsymbol{L}_{\boldsymbol{k}}$	Study covariates at time k**	Employment status at age 19, Job title in 1930
$\bar{\boldsymbol{L}}_{\boldsymbol{k}}$	Covariate histories at time k**	Employment duration at age 19, Number of jobs worked by 1930
$\boldsymbol{V}$	Study covariates fixed in time within the study†	Age at hire, birth cohort, race
T	Time to event of interest	Age at death from lung cancer
$\boldsymbol{U}$	Unmeasured variables	Underlying health status, frailty, smoking (if unmeasured)

Note: all subscripts denoting individuals are suppressed for clarity
* Bolded variables refer to a possible vector (e.g. the collection of race, birth cohort, and age at hire), non-bold variables refer to a scalar (e.g. cumulative exposure at age 40), while lowercase variables refer to a realization (e.g. 10 WLM of exposure)
** Also referred to as time-varying covariate
† Also referred to as a time-fixed covariate

1.2.1.2 A model for healthy worker survivor bias based using directed acyclic graphs

Robins (1986) proposed that healthy worker survivor bias could be (partially) explained by a general set of causal mechanisms in occupational studies and showed that, under this assumed model, that conventional statistical methods cannot adequately reduce the amount of bias that results from these mechanisms. He posited that healthy worker survivor bias occurs when:

1. Employment status (whether or not at work) is an independent population risk factor for a disease outcome of interest, possibly because of association through an unmeasured health determinant that affects both employment and the outcome U.

2. Previous employment status affects subsequent exposure (for example, if exposure only happens at work)

3. Previous exposures impact the rate of leaving work

Using the language of directed acyclic graphs (hereafter causal diagrams), we can formally express these relationships using simple graphs (*Pearl* (1995); *Greenland et al.* (1999)). These diagrams allow a simple assessment of conditional dependence of a set of variables of interest - typically these variables are the exposure of interest, the outcome of interest, and all known confounders of the exposure-outcome association. The causal diagrams are formal representations of graph-theory, which allows a formal assessment of possible structural biases (*Greenland and Pearl* (2008)) and can be extended to other situations such as bias due to missing data (*Daniel et al.* (2011)).

When one has some idea about causal relationships in the data, causal diagrams are a useful tool for selecting an analytic plan. Causal diagrams, for example, are useful in occupational studies for determining when regression based methods can be used to control confounding bias (*Buckley et al.* (2014)). As an example, Figure 1.2 shows

a possible mechanism underlying healthy worker survivor bias proposed by *Robins* (1986).

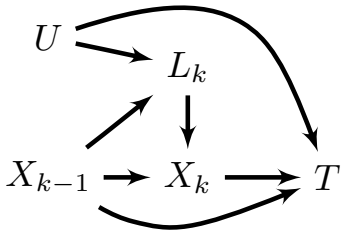

Figure 1.2: Causal diagram showing causal relationships possibly underlying healthy worker survivor bias in a hypothetical occupational studies with notation given in Table 1.2.

This diagram can be used to show that regression based methods will be biased in this scenario. Using the variables shown in figure 1.2, consider a proportional hazards model of the form

$$\lambda(k|X_k, X_{k-1}, L_k) = \lambda_0(k)\exp(\beta_1 X_k + \beta_2 X_{k-1} + \beta_3 L_k)$$

Where $\lambda(k|X_k, X_{k-1}, L_k)$ is the hazard at time k conditional on X_k, X_{k-1}, L_k, and $\lambda_0(t)$ is an unspecified baseline hazard for individuals at the referent level of each variable. The vector of β parameters represent the log hazard ratio for a one unit change in each variable, while holding the other variables constant.

We are interested in estimating the effects of exposure $\bar{\boldsymbol{X}}_{\boldsymbol{k}} \equiv (X_k, X_{k-1})$ on the hazard or time to death. That implies we are primarily interested in β_1 and β_2. The diagram shows that there is confounding of the $X_k \rightarrow T$ association along the pathway $X_k \leftarrow L_k \rightarrow U \rightarrow T$. This is how the healthy worker survivor bias can be conceptualized: confounding that occurs because a common factor causes both attrition from the workforce and the outcome of interest. This is often referred to as time-varying, or time-dependent confounding (*Robins* (1992); *Daniel et al.* (2013)).

In our model, we can control this confounding by including L_k in the model. However, including employment status would cause bias in the coefficient β_2. The variable

L_k is referred to as a collider on the path between X_{k-1} and k because, along the path $X_{k-1} \rightarrow L_k \leftarrow U \rightarrow T$, two arrow heads "collide" at L_k. Adjusting for L_k in this model will up a non-causal, backdoor path from exposure to outcome, a situation known as collider bias (*Cole and Hernán* (2002); *Greenland* (2003); *Cole et al.* (2010)).

To see this bias, consider an individual at time k. Prior exposure and poor health can both cause individuals to leave work, so if that individual is off work, we know he either likely to be highly exposed or of poor baseline health. If he is unemployed and highly exposed, he is more likely to have good health and less likely to suffer the outcome of interest. Thus, within strata of employment, exposure and the outcome are associated, even if there is no effect of exposure on the outcome.

More relevantly, collider bias will still occur if we consider cumulative exposure $(X_{k-1} + X_k)$ as the exposure of interest (or any summary of both time-specific exposures). That is equivalent to our model above if we add the restriction that $\beta_1 = \beta_2$. Because we know that β_2 is biased if we adjust for L_k, and β_1 is biased if we do not adjust for L_k, the association between cumulative exposure and the outcome will also be biased . Note that, had we measured U, we could control this confounding by adjusting for it in the model, instead of L_k. Given the limited nature of occupational data, which is often limited to employment records, it is unlikely that we could adequately capture U in most circumstances.

Bias would also occur if employment status directly affected the outcome (figure 1.3, say, by an increase in smoking after leaving a non-smoking workplace, or a loss of insurance benefits), and bias would be present even if exposure did not have an effect on the outcome (*Rosenbaum* (1984)). This bias occurs because L_k is an intermediate between X_{k-1} and the outcome, so adjusting for it would be equivalent controlling some of the effect of exposure.

In Robins' model for healthy worker survivor bias, it is worth highlighting that there are two central biases at play: 1) not accounting for employment status as a time vary-

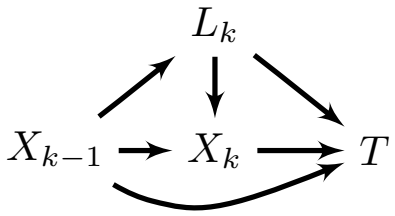

Figure 1.3: Causal diagram showing causal alternative relationships possibly underlying healthy worker survivor bias with notation given in Table 1.2. Employment directly affects the outcome of interest.

ing confounder, or 2) bias from improperly accounting for employment status. The observed bias when considering cumulative exposure is generally downward because low cumulative exposures result from individuals leaving work who have a poor prognosis. Adjustment through conventional means may also yield a downward bias, however, since it is conditioning out part of the total effect of exposure. Thus, based only on the simple model shown in figure 1.3 one might conceivably adjust for employment status in a regression model and see no change in the effect estimate because one bias has been traded for another.

Note that, if employment status is not affected by prior exposure (i.e. only 1 and 2 from Robins' conditions for healthy worker survivor bias hold), then less restrictive methods could be used to adjust for time-varying confounding by employment status (*Pearce* (1992); *Steenland and Stayner* (1991)). Further, *Pearce et al.* (2007) notes that, along with employment status, job-title or work area can also function as time-varying covariates that function like employment status. That is, exposure can cause irritation, leading to a transfer of positions, or, alternatively, sensitive individuals can also be transferred to low-exposure jobs - this leads to a similar selection process to that posited by Robins. In the nuclear industry, if an employee reaches his or her exposure limit, a job transfer to a less exposed job (or termination) could occur, suggesting that exposure could be a strong risk factor for job changes or employment status changes.

1.2.2 Evidence of a possible healthy worker survivor bias in occupational radon studies

Healthy worker survivor bias has historically been of concern only when it results in benign or deleterious exposures appearing to be beneficial. For example, in a cohort of Swedish iron-ore miners, *Björ et al.* (2013) observed negative trends in multiple outcomes with increasing employment time and employment time underground, as shown in Figure 1.4. These trends occurred even though the miners are exposed to radon, silica, and diesel exhaust, and dust in the underground environment - previous analyses found associations between cumulative radon exposure and lung cancer (*Jonsson et al.* (2010)). The negative association between employment time and rectal cancer may be present due to beneficial effects of the physical activity inherent in mining (*Slattery* (2004)), but the positive effect of physical activity is not strong enough to induce such associations. Thus, there is evidence of healthy worker survivor bias in a cohort of radon exposed miners, but it may not be apparent with respect to lung cancer because radon has highly specific effects and the effects are large enough to outweigh any downward bias. This bias may be of little concern when testing hypotheses about highly deleterious exposures. However, when one is interested in characterizing a dose-response, healthy worker survivor bias should be considered a potential problem even if it does not completely eliminate apparent associations.

1.2.2.1 Health related selection is observable in miner cohorts

Evidence of the healthy hire effect in miner cohorts may provide indirect evidence of a healthy worker survivor bias. Presence of a healthy hire effect suggests that employment depends on some underlying health status. As shown in the simple causal diagram in Figure 1.5, it is apparent that healthy worker survivor bias and the healthy hire effect differ by only one arrow on the diagram. In this scenario, we define $k-1$ as the time at which the cohort was formed, so L_{k-1} represents initial employment in

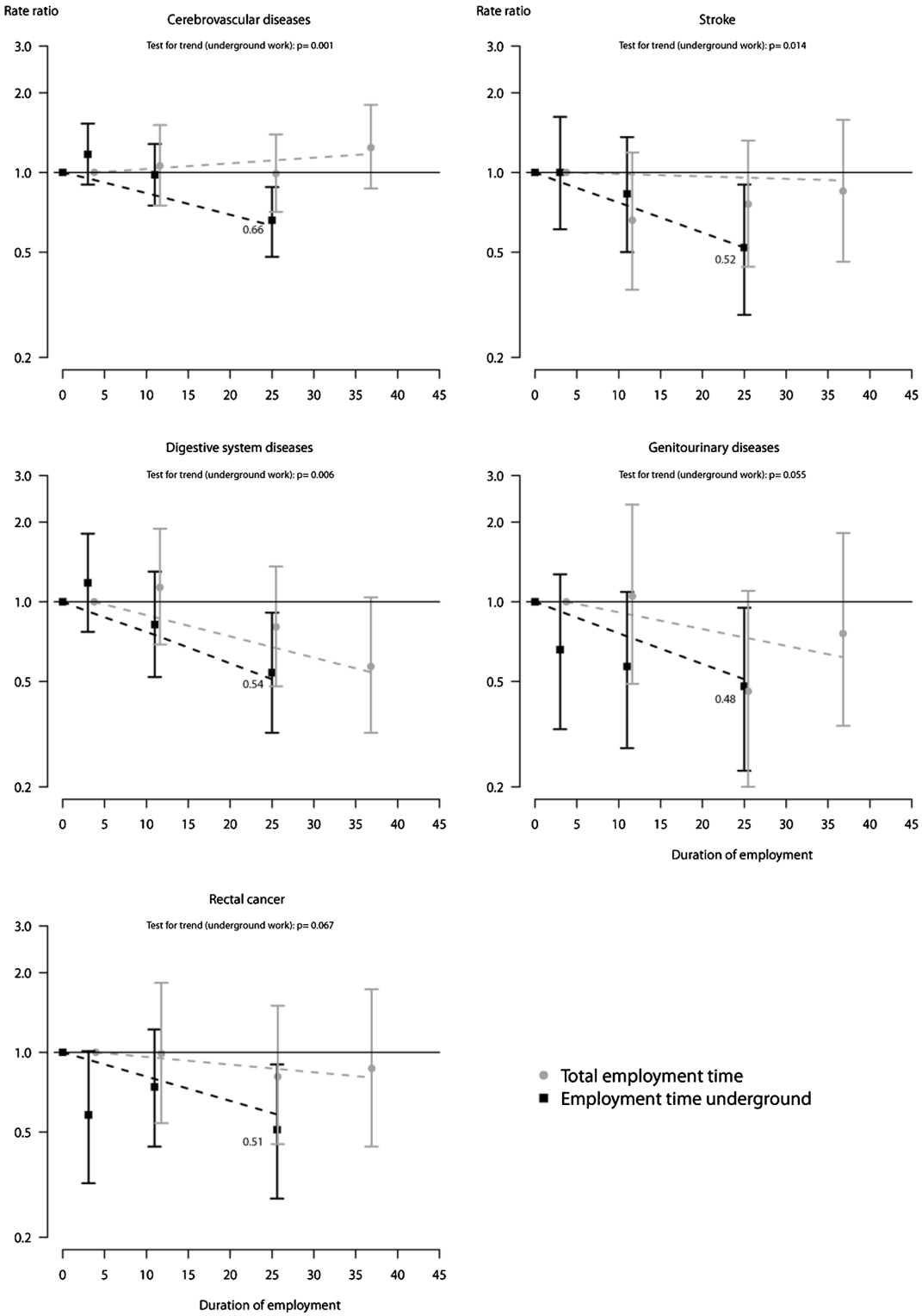

Figure 1.4: Figure 1 from *Björ et al.* (2013) showing negative trends in multiple outcomes with employment duration in an iron mine

the industry under study. The dashed arrow in Figure 1.5 represents the difference between these two biases. Healthy worker survivor bias suggests that underlying health status U affects employment status at all time points (i.e. the dashed arrow in figure 1.5 is present). When the healthy hire effect (and not healthy worker survivor bias) is present, then U affects employment status only at time $k-1$ (the dashed arrow is absent). This diagram is an obvious over-simplification of the complex forces that drive employment changes across time. However, it highlights the similarity of the two sources of bias in occupational studies and clarifies the specific conditions under which one could observe a healthy hire effect but expect no healthy worker survivor bias.

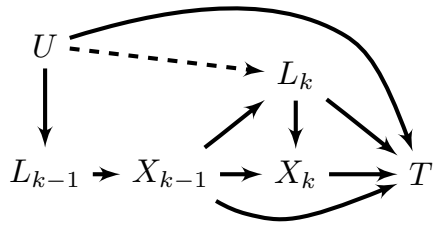

Figure 1.5: Directed acyclic graphs comparing a model for healthy worker survivor bias (dashed arrow is present) with a model for the healthy hire effect (dashed arrow is absent)

Often, standardized mortality ratios (SMRs) are interpreted as evidence of the healthy hire effect if they are below one. Deleterious effects of occupational exposures may counterbalance this effect, and strongly deleterious exposures may cause an excess of mortality in working populations. In an early analysis of the Colorado Plateau uranium miners, *Archer et al.* (1976) estimated an all cause SMR of 0.86 for American Indians, with an SMR for heart disease estimated to be 0.10, and a deficit of mortality from non-accidental, non-respiratory causes among white miners. *Schubauer-Berigan et al.* (2009) estimated an all cause (except for lung cancer) SMR of 1.22, while the SMR for ischemic heart disease was 1.05, suggesting that the healthy hire effect has not persisted in this cohort. Among a cohort of Canadian miners, *Howe et al.* (1988) noted deficits

of circulatory diseases (SMR=0.65), non-lung cancer respiratory diseases (SMR=0.75), and non-respiratory cancers (SMR=0.80), while the all cause mortality SMR of 1.03 suggests that occupationally related diseases (e.g. accidents, SMR=1.86 and lung cancer, SMR=1.90) may be masking a healthy hire effect in this population. The low SMR for non-lung cancer respiratory diseases remained in deficit in the most recent update of this population (SMR=0.80) while the all-cause mortality SMR remained near unity (SMR=0.97) (*Lane et al.* (2010)).

Stayner et al. (2003) attributes non-linearity observed Colorado uranium miners (see *Hornung and Meinhardt* (1987)) and in Czechoslovakian uranium miners (see *Sevc et al.* (1993)) to a lower than expected risk at high exposure levels. Because the phenomenon is not generally observed in miner studies (see *NRC* (1999)), *Stayner et al.* (2003) posited that non-linearity results from particular aspects of the cohorts under study, such as measurement error or healthy worker survivor bias. In an analysis of non-respiratory cancers in 11 underground miner cohorts, *Darby et al.* (1995) noted that, both within each cohort and in the joint analysis, SMRs were nearly all higher for workers that had been employed at least 10 years versus those employed fewer than 10 years, indicating some health related selection. *NRC* (1999) have noted the non-linearity at high exposures but addressed it by truncating person-time-at-risk when individuals reach the highest exposure levels. Thus, there is reason to suspect that heterogeneity in the ERR/WLM from the miner studies may result partially from healthy worker survivor bias, which may be stronger in some cohorts.

1.3 Models of the dose-response relationship between radon and lung cancer

As we postulate in §1.2.2, there may be healthy worker survivor bias in studies of radon exposure in miners. However, as we noted in §1.2.1.2, this bias may not be reduced by controlling for time-varying confounding by employment status in a regression model. We review some of the recent models used in dose-response analyses for

the radon-lung cancer dose-response and comment on the implications for healthy worker survivor bias.

Case reports of respiratory diseases in miners date back to Paracelcus (see *Jacobi* (1993) for a history). However, uranium miner cohort studies did not begin in earnest until the 1960s. Since that time the radon-lung cancer dose-response has been extensively studied. Recent efforts have focused on estimating a dose-response at the "low" radon concentrations observed in residential settings using meta-analysis or pooled data from miners studies (e.g. *Lubin* (1994); *Lubin et al.* (1995a, 1997); *Leuraud et al.* (2011)), residential studies (e.g. *Lubin and Boice* (1997); *Lubin* (2003); *Krewski et al.* (2005, 2006)), a combination (e.g. *Fornalski and Dobrzynski* (2011)), or specific subgroups of miners or residential populations (e.g. *Lubin et al.* (1994)).

Much work has been put into creating parsimonious models of the radon-lung cancer dose response that follow basic principles. One of these principles, the *linear-no-threshold* (LNT) assumption, is common to many radiation studies. This assumption implies that the dose-response between radon and health outcomes will be linear and that exposure to radon at any level will increase the risk of adverse health outcomes (see *Brenner and Sachs* (2006) for review of the mechanistic basis for the LNT).

The Committee on the Biological Effects of Ionizing Radiation VI (BEIR VI) (based on models developed by *Lubin* (1994)), chose to model the radon-lung cancer association using a linear excess relative rate (ERR) model, shown in equation 1.1.

$$\lambda(k|\bar{X}_{k-5};\beta) = \lambda_0(k)(1+\beta\,\bar{X}_{k-5}) \tag{1.1}$$

An ERR model says that the rate of disease $\lambda(k)$ increases linearly with the exposure of interest - here expressed as cumulative radon with a 5 year lag $\bar{X}_{k-5}$ and β represents the ERR per unit of additional exposure, given a baseline rate of disease $\lambda_0(k)$. The baseline rate may be specified as a parametric distribution, such as the exponential

distribution, or it may be left unspecified as in a semi-parametric Cox proportional hazards model. This model can be expressed equivalently as

$$ERR(\bar{X}_{k-5};\beta) = \beta \bar{X}_{k-5} \tag{1.2}$$

For risk projections, the authors of the BEIR VI report considered categorical versions of both the "exposure-age-duration" and "exposure-age-concentration" models (known collectively as the BEIR VI models *NRC* (1999), p. 80). The BEIR VI models allowed for the ERR to vary over categories of time since exposure attained age and either exposure duration (cumulative up to time k) or exposure concentration (average up to time k) as in

$$ERR(\bar{\boldsymbol{X}}_{\boldsymbol{k-5}}\bar{L}_{k1},\bar{L}_{k2};\beta,\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\gamma}) = \beta(\bar{X}_{k-[5,15)} + \theta_1\bar{X}_{k-[15,25)} + \theta_2\bar{X}_{k-25}) \times \phi L_{k1} \times \gamma L_{k2} \tag{1.3}$$

where $\bar{\boldsymbol{X}}_{\boldsymbol{k-5}} = (X_{k-[5,15)}, X_{k-[15,25)}, X_{k-[25+)})$ is the exposure to radon in working level months for the period 5-15 years, 15-25, or 25+ years years prior. There is modification of the ERR by attained age $\bar{L}_{k1}$, and $\bar{L}_{k2}$ can be either exposure rate (in working levels) or exposure duration. The coefficient β is the ERR/WLM at the reference level of all modifiers, the θ parameters represent the factor by which the ERR/WLM is multiplied for the time-window specific exposure, and ϕ and γ are the factors by which the ERR/WLM is multiplied for their respective covariate. This model is equivalent to fitting a model with the three windows of exposure within strata of each of the time-varying factors, but the model given in gives explicit parameters for the strength of effect measure modification by the time varying factors. Note that the model assumes that exposures in the previous five years have no effect on lung cancer outcomes (i.e. they assumed a five year lag).

Many recent analyses in individual miner cohorts utilize models similar to the BEIR VI models. In an analysis of German uranium miners, *Kreuzer et al.* (2010) used a model

which differs from the BEIR VI models only in that time since exposure is not considered to modify the relative risk of lung cancer by radon exposure.

While the BEIR VI models include factors that multiply the ERR on a linear scale, there is no restriction on the modification. For example, in the model shown in 1.3, there is no restriction on ϕ, so strong modification by exposure rate could lead to the estimate of a negative ERR.

To get around this issue, other models were developed in which which modifying factors (e.g. time since exposure) are considered to have a log-linear relationship with the relative risk of lung cancer. For example, in a cohort of Canadian uranium workers, *Lane et al.* (2010) and in a joint analysis of French and Czech miners *Tomásek et al.* (2008) considered (in addition to the BEIR VI "exposure-age-concentration" model), models of the general form

$$RR(\bar{\boldsymbol{X}}_{\mathbf{5}}, \bar{\boldsymbol{L}}_{\boldsymbol{k}}; \beta, \gamma) = 1 + \bar{\boldsymbol{X}}_{\mathbf{5}} \beta \exp\left(\bar{\boldsymbol{L}}_{\boldsymbol{k}} \gamma\right) \tag{1.4}$$

This model allows modifiers to enter the model as linear terms (rather than categorical as in the BEIR models) - otherwise, a linear trend may predict a negative ERR for certain combinations of exposure and the modifiers.

Leuraud et al. (2011) considered a model that further generalizes that in 1.4 by allowing for modification of the baseline rate (at the reference level for all exposures and modifiers) by fitting a model of the type

$$RR(\bar{\boldsymbol{X}}_{\mathbf{5}}, \bar{\boldsymbol{L}}_{\boldsymbol{k}}, S_k; \beta, \gamma, \theta) = \exp(S_k \theta)\left[1 + \bar{\boldsymbol{X}}_{\mathbf{5}} \beta \exp\left(\bar{\boldsymbol{L}}_{\boldsymbol{k}} \gamma\right)\right] \tag{1.5}$$

Where S_k in the study was current smoking status. Note that for $\beta = 0$ this corresponds to a standard log-linear rate model.

Each of the models given in 1.3-1.5 is an attempt at capturing important aspects of exposure history that predict subsequent risk better than cumulative exposure, as in

the simple model given in 1.2. An equivalent approach has been applied to the relationship between smoking and lung cancer (*Vlaanderen et al.* (2014)) as an attempt to improve on the pack-years metric (*Thomas* (2014)), which is the analog of the cumulative exposure used in the radon literature.

A shortcoming of each of the models 1.3-1.5 is that the modifiers they consider are restricted to modification by historical summaries or current values of the modifiers. Generally, historical summaries are needed for factors such as exposure rate (e.g. average exposure rate up to age k) and age at exposure (e.g. age at median exposure) because mortality generally does not occur until long after exposure has ceased.

As a more general approach to modeling the time-related aspects of exposure history, *Richardson et al.* (2012) considered an approach that was not a simple variation on the models 1.3-1.5. Rather than considering how time related aspects of exposure might modify the effects of cumulative exposure, the authors considered how these factors might modify the cumulative exposure itself. In this model, given in 1.6, the summary exposure metric changes with modifying factors. The summary metric can be considered a time varying weighted sum of exposure $\sum_{u=0}^{k} \bar{\boldsymbol{X}}_{\boldsymbol{u}} w_u$, such that when the weight $\bar{\boldsymbol{w}}_{\boldsymbol{k}} = (w_0, ..., w_k)$ is always one, this summary metric is equal to the cumulative exposure. Otherwise, this weight varies as a function of the collection of modifiers included in $\bar{\boldsymbol{L}}_{\boldsymbol{k}}$ and the parameter vector γ, which quantifies the strength of this modification.

$$ERR(\bar{\boldsymbol{X}}_{\boldsymbol{k}}, \bar{\boldsymbol{L}}_{\boldsymbol{k}}; \beta, \gamma) = 1 + \sum_{u=0}^{k} \bar{\boldsymbol{X}}_{\boldsymbol{u}} \beta \exp\left(\bar{\boldsymbol{L}}_{\boldsymbol{u}} \gamma\right) \tag{1.6}$$

Summary, and implications for healthy worker survivor bias The fields of radiation and occupational epidemiology have been concerned mainly about the modification of the effect of radon over time. The two approaches consider modification of the association between cumulative exposure and lung cancer or weighting of the relevant aspect of the exposure metric. Notably, all of these models are unified in that they

consider some time-varying summary of exposure history to be the relevant predictor of lung cancer mortality, rather than, say only current exposure. When some of these potential modifiers may also be confounders, however, this set of approaches may be biased. As shown in §1.2.1.2, if exposure at prior time points can affect the subsequent evolution of these modifiers, then bias will result. For example, if prior exposure influences the rate of leaving employment, then models that do not stratify by some function of employment history may be biased due to confounding. However, models that stratify by such factors - including exposure duration, which is strongly affected by employment history - will also be biased. Thus, the possible presence of healthy worker survivor bias in studies of radon exposed populations motivates the need for models that can account for a time-evolving exposure while not being subject to the shortcomings of regression models. We explore such models in the next section.

1.4 Existing analytic methods proposed for reducing bias due to healthy worker survivor bias

1.4.1 Parametric g-computation algorithm

Robins formulated the concept of healthy worker survivor bias as a case of time-varying confounding in which (some) of the confounders also mediate the exposure-disease relationships (*Robins* (1987b,a, 1986); *Robins et al.* (2004)). In his original formulation, he proposed use of the parametric g-computation algorithm to identify causal effects in cohorts in there is healthy worker survivor bias. The parametric g-computation algorithm, also known in various formulations as g-computation, the g-formula, parametric g-formula (*Taubman et al.* (2009)), and model-assisted g-computation algorithm (*Robins* (1989)), is based on standardization. A simple example of a standardized (marginal) probability that $A = 1$, given a binary variable B can be expressed as:

$$Pr(A=1)=\sum_{b\in(0,1)} Pr(A=1|B=b)\times Pr(B=b) \tag{1.7}$$

Unlike the representation of the ERR model in equation 1.1, this way of expressing standardization requires use of discrete time - and slightly different notation. In this case, the outcome of interest is an outcome D_k that is measured at discrete points in time, such as a study visit. We define the g-formula using a simple data structure in which, at time k exposure ($\boldsymbol{X_k}$), employment status ($\boldsymbol{L_k}$), and the death (D_k) are all dichotomous variables. The cumulative incidence, I_k is the proportion of individuals who have died by time k. To estimate I_k we can factor $Pr(D_k=1)$ as in 1.7. The standardized (marginal) cumulative incidence (I_k) of the dichotomous outcome D_k (assuming no censoring) can be expressed using the g-formula as:

$$\begin{aligned} I_k = Pr(D_k=1) = \\ &\sum_{j=1}^{k}\sum_{l}\sum_{x} Pr(D_j=1|\bar{\boldsymbol{X}}_{\boldsymbol{j}}=\bar{\boldsymbol{x}}_{\boldsymbol{j}},\bar{\boldsymbol{L}}_{\boldsymbol{j}}=\bar{\boldsymbol{L}}_{\boldsymbol{j}},D_{j-1}=0)\times \\ &\prod_{m=1}^{j}[Pr(X_m=x_m|\bar{\boldsymbol{L}}_{\boldsymbol{m}}=\bar{\boldsymbol{l}}_{\boldsymbol{m}},\bar{\boldsymbol{X}}_{\boldsymbol{m-1}}=\bar{\boldsymbol{x}}_{\boldsymbol{m-1}},D_{m-1}=0)\times \\ &Pr(L_m=l_m|\bar{\boldsymbol{L}}_{\boldsymbol{m-1}}=\bar{\boldsymbol{l}}_{\boldsymbol{m-1}},\bar{\boldsymbol{X}}_{\boldsymbol{m-1}}=\bar{\boldsymbol{x}}_{\boldsymbol{m-1}},D_{m-1}=0)\times \\ &Pr(D_{m-1}=0|\bar{\boldsymbol{L}}_{\boldsymbol{m-1}}=\bar{\boldsymbol{l}}_{\boldsymbol{m-1}},\bar{\boldsymbol{X}}_{\boldsymbol{m-1}}=\bar{\boldsymbol{x}}_{\boldsymbol{m-1}},D_{m-2}=0)] \end{aligned} \tag{1.8}$$

Where variables with subscripts ≤ 0 drop out if $m <= 2$.

The g-formula can be used for effect estimation by estimating the cumulative incidence under different interventions on exposure. For example, $I_k^{\bar{\boldsymbol{x}}_k}$ is the cumulative incidence at time k we would expect under the intervention $set(\bar{\boldsymbol{X}}_{\boldsymbol{k}}=\bar{\boldsymbol{x}}_{\boldsymbol{k}})$. The cumulative incidence under an intervention is estimated using the g-formula by replacing $\bar{\boldsymbol{x}}_{\boldsymbol{k}}$ with the value to which we wish to set exposure. For example, setting the exposure to always equal zero would would be the intervention "never exposed." To estimate the

risk difference at time k for always versus never exposed, we can take the difference between two interventions $I_k^{\bar{1}}$-$I_k^{\bar{0}}$.

Non-dichotomous data, moderate to small sample size and extended follow up (i.e. any condition that results in sparse strata) necessitate use of models and Monte Carlo sampling. This approach is referred to as the parametric g-formula (see examples in *Robins et al.* (2004); *Taubman et al.* (2009); *Westreich et al.* (2012); *Cole et al.* (2013); *Keil et al.* (2014a)). Recently, *Edwards et al.* (2014) used the g-formula to estimate the effects on mortality of implementing and enforcing a series of different occupational limits on exposure in the study population used in the analyses in chapters 3.1 and 3.2, the Colorado Plateau uranium miners.

We note that the g-methods (the parametric g-formula, g-estimation of structural nested models, inverse probability weighted marginal structural models) originate in the causal inference literature and under the assumptions given in Appendix C.3 the parameters from these methods may have a causal interpretation. Without necessarily assuming we have identified a causal effect, however, g-methods have some advantages over standard regression models. Under the conditions for healthy worker survivor bias outlined by *Robins* (1986), g-methods appropriately adjust for time-varying confounding in which some of the confounders are also intermediates. Namely, as shown with the g-formula examples, the parameter of interest is the marginal incidence under an intervention $I_k^{\bar{x}_k}$. This marginal incidence is not stratified by the time-varying confounder, so it does not fall prey to the pitfalls of stratifying or conditioning on a variable affected by exposure that we noted in §1.2.1.2 and were discussed by *Rosenbaum* (1984) and *Weinberg* (1993).

1.4.2 Other proposed solutions

Arrighi and Hertz-Picciotto (1993, 1994, 1996) review several other methods for controlling bias due to healthy worker survivor bias, and *Arrighi and Hertz-Picciotto* (1996)

compare them in an analysis of the association between arsenic exposure and cancer outcomes in a cohort of copper smelters, while *Hertz-Picciotto et al.* (2000) compare the methods using the same cohort in an analysis of the arsenic-cardiovascular disease association. Three methods were compared with the g-null test (an early g-method similar to the parametric g-computation algorithm - see *Robins* (1987b)): exposure lag, adjustment for employment status, cohort restriction. Using causal diagrams we can show the specific conditions under which these analyses may control healthy worker survivor bias.

Exposure lag *Gilbert* (1982) employed use of an exposure lag to control healthy worker survivor bias under the rationale that the most recent exposures could only be accrued by employees that survived/stayed at work (namely, the healthiest employees). Under figure 1.6, this is equivalent to removing the arrow $X_k \rightarrow \mathrm{T}$. Doing so, and not including X_k in the model for lung cancer, ensures that exposure is not confounded by L_k, since L_k only confounds the $X_k \rightarrow \mathrm{T}$ association. If there is, in reality, a $X_k \rightarrow \mathrm{T}$ association, then there will be induction of some measurement error due to the fact that were are ignoring X_k. If we consider a more realistic setting, then there will be a likely trade-off between measurement error due to ignored exposure and confounding by employment status that depends on the length of lag used.

In an analysis of the arsenic-lung cancer association, *Arrighi and Hertz-Picciotto* (1996) observed increased rate ratios for lags of 10 and 15 years (versus no lag) across multiple exposure levels, but a lag of 20 years produced rate ratios that were both more variable and generally lower than the other lags. These findings suggest some trade-off may be occurring between reduction of confounding due to healthy worker survivor bias and some increase in the variance of the rate ratio estimate (and reduction of the effect estimate, in the extreme), which is consistent with measurement error (that is, error in measuring etiologically relevant exposures) introduced due to lagging. It

should be noted that exposure lagging in this context is identical to exposure lagging to account for the empirical latency period between exposure and disease onset *Rothman* (1981); *Richardson et al.* (2011). The results of *Arrighi and Hertz-Picciotto* (1996) could also be a consequence of modeling the disease process more effectively in models using a lag that approaches the average empirical latency period in the cohort.

In a simple model of healthy worker survivor bias shown in figure 1.6, utilizing a lag of one unit of time would appropriately adjust for confounding by employment status without inducing additional bias. Using this diagram, confounding occurs along $X_k \leftarrow L_k \leftarrow U \rightarrow T$, and by excluding X_k (through the lag), we have eliminated confounding of the effect of the exposure of interest $\bar{X}_k = (X_{k-1}, X_k)$. This model requires that there is no time-varying confounding of $X_{k-1} \rightarrow T$ and is not likely to hold in any realistic setting. Note that lags may still be appropriate to address disease latency, but analysis of our causal diagram in Figure 1.6 suggests that exposure lagging will not be sufficient to control healthy worker survivor bias. However, it may be useful in concert with other methods (e.g. *Garshick et al.* (2012),*Naimi et al.* (2014a)).

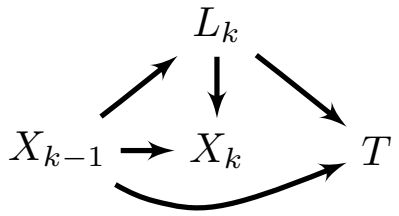

Figure 1.6: Causal diagram showing assumption under which a 1 unit exposure lag controls confounding by employment status

Adjustment for employment status *Gilbert and Marks* (1979) proposed that stratifying person-time by current employment status (on versus off work) could be used to control aspects of bias due healthy worker survivor bias. The relationships shown in Figures 1.2-1.3 again suggest scenarios in which employment status stratification (adjustment) could reduce this bias. Employment status at time k lies on a non-causal,

backdoor path between exposure at k and the outcome. Stratification on current employment status would block confounding pathways for X_t, but would also block part of the total effect of exposure at time 1, thus biasing the estimate of total exposure. Additionally, this adjustment induces collider-stratification-bias by opening up a non-causal, backdoor path $X_{k-1} \rightarrow \not{L_k} \leftarrow U \rightarrow T$, where $\not{L_k}$ means that the model was stratified by L_k (*Greenland* (2003); *Cole et al.* (2010)). If exposure does not affect subsequent employment status - i.e. if exposure has no effect on health during employment, then stratification of person time would function to eliminate confounding by employment status (e.g. the causal diagram shown in figure 1.3). If the confounding through L_k is strong relative to the indirect effect of prior exposure (X_{k-1}) through L_k, then adjustment may increase effect estimates. *Steenland and Stayner* (1991) observed such an increase and concluded that employment status is an important variable for adjustment in occupational studies. Such an increase does not rule out collider bias, however.

A similar approach was proposed by *Flanders et al.* (1993), who explored an empirical model for the healthy hire effect in which they proposed adjustment for time-since-hire to reduces bias due to healthy worker effects. The authors also posited that the healthy hire effect faded over time through a mechanism in which the hazard increased in proportion to the number of years employed, suggesting that cumulative work (conditional on age) actually makes workers more sick. Interestingly, the authors observed a downward bias due in risk ratios (compared to employment duration adjusted models) to their hypothesized set of effects, which is same as that seen by *Steenland and Stayner* (1991) among several work forces. *Cardis et al.* (2007) observed a 0.31 ERR/Sv for radiation dose effects on all cancers (excl. leukemia) before adjustment by duration of employment and a 0.97 ERR/Sv after adjustment in the 15 country study of radiation exposed nuclear workers. These results conform to the theoretical direction of bias due to healthy worker survivor bias in work by *McNamee* (2003) and *Robins* (1986). *Flanders et al.* (1993) posited that the bias due to healthy worker survivor bias

could be eliminated by adjusting for time-since-hire, but *Arrighi and Hertz-Picciotto* (1995) suggests using results from two cohorts that the empirical model of *Flanders et al.* (1993) does not hold due to apparent confounding of the time-since-hire effect by age and employment status.

Richardson et al. (2004) showed in simulations that bias patterns observed in healthy worker survivor bias could result from a strong correlation between cumulative exposure and time-since termination. The authors noted in simulations that large, spurious associations could be induced by simply adjusting for a binary indicator of employment status, whereas adjustment for time-since-termination returned the expected result. The methods were replicated in a cohort of utility workers with the same pattern of results, and later applied by *Garshick et al.* (2008) to a cohort of trucking industry workers. *Steenland and Stayner* (1991) observed that the SMR during inactive (unemployed) person time was higher than that during employment, and a decrease in the SMR by duration of employment - the confluence of which is consistent with sick workers leaving, rather than employment increasing the health of employees. These results emphasize the importance of considering employment status as a time-varying factor in any occupational study, as well as considering a set of employment status "history" variables that could include time-since termination or duration of employment.

Cohort restriction Following observations that SMRs for multiple diseases among vinyl-chloride exposed workers with 15+ years at work were elevated versus those with 5+ or 10+ years at work, (*Fox and Collier* (1976, 1977)) Fox and Collier advocated restriction of occupational cohorts to those with at least 15 years of employment. The authors attributed this shift to health related selection from the workforce that decreased after time spent in the industry. *Arrighi and Hertz-Picciotto* (1996) also explored cohort restriction in a cohort of copper smelters exposed to arsenic and found elevated rate ratios for respiratory in cohorts restricted to 20+ years of work, but not in those

restricted to 10+ or 15+.

In the causal diagram in Figure 1.6, this method can be expressed as restricting the cohort to those with $L_2 = 1$. This restriction is fraught with the same issues as regression adjustment for L_2. Further, there is no reason to suspect that selection effects from healthy worker survivor bias are decreased at all in restricted cohorts, and it is impossible to disentangle possible control of any healthy worker effects from the effects of selection bias induced by restricting analysis cohort that has a) survived long enough to be in the cohort and b) been subjected to more exposure than others.

The g-null test The g-null test was another approach developed by *Robins* (1986) to address healthy worker survivor bias when exposure may influence employment. This method is similar to g-estimation of structural nested models (discussed in §1.5.1), but is less general. The approach relies on creating nested matched sets individuals, where the matching is on exposure and covariate history. Each nested set is analyzed with the equivalent of a conditional logistic model where a case is defined as an individual who eventually dies of the disease of interest. The test was low powered, so *Robins* (1987b) developed a slightly more powerful version that was applied by *Hertz-Picciotto et al.* (2000) to study the effects of arsenic on circulatory disease. While the approach might be tenable for short term studies with only categorical covariates, it appears that g-estimation of structural nested models may offer the same advantages without any of the disadvantages of the g-null test.

Summary Robins' model for healthy worker survivor bias is inclusive of many previously suggested models (e.g. *McNamee* (2003); *Flanders et al.* (1993); *Pearce* (1992); *Richardson et al.* (2004)). The causal diagrams in figures 1.2-1.6 are all examples of causal structures for which exposure effects can be appropriately estimated without inducing other biases (*Robins and Wasserman* (1997)). Doing so requires using g-methods, such as the parametric g-formula, g-estimation of structural nested models, or inverse

probability weighted marginal structural models. In Figures 1.2-1.5, if exposure does not influence subsequent employment, then adjustment for employment status would also be expected to remove confounding without introducing additional bias. Ultimately, these simplified scenarios are useful for suggesting approaches in certain scenarios, but the often unobservable complexity of relationships between the external environment and health implies that we can only approximate the truth with models. The flexibility of g-methods makes them a useful tool to in a wider array of environments than standard regression models.

1.5 Proposed solutions considered in this dissertation

Two methods aside from the g-formula designed for the study of exposure effects under time varying confounding are inverse probability of exposure weighted marginal structural models (MSM), and g-estimated structural nested models. Of these, inverse probability weighted Cox MSMs and structural nested accelerated failure-time models (SNAFT model) have utility for data in which the effect of cumulative exposures on a binary outcome are of interest. Both models can potentially be used to control time varying confounding by employment status and employment history, which We hypothesize will control healthy worker survivor bias. We propose to apply these two methods, and We describe simple examples of them in §1.5.1 and §1.5.2.

1.5.1 Structural Nested Accelerated Failure Time models

Robins proposed the use of structural nested models (specifically Structural Nested Accelerated Failure Time [SNAFT] models in the case of survival outcomes) as a way to estimate causal effects in the presence of time-varying mediating confounders (*Robins and Tsiatis* (1992)). These models adjust for confounding by time-varying covariates using g-estimation (*Robins* (1989)). G-estimation allows estimation of a marginal pa-

rameter, as with the parametric g-formula, and is not subject to the pitfalls of regression models noted above.

A SNAFT model can be expressed in a simple case as

$$T^{\bar{\mathbf{0}}} = \int_{k=0}^{T} \exp(\psi X_k) d\,k \tag{1.9}$$

Where $T^{\bar{\mathbf{0}}}$ is the failure time we would observe if the individual had never been exposed. The parameter of interest $\exp(-\psi)$ is interpreted as the factor by which exposure contracts one's lifespan, For example, if $\exp(-\psi) = 1/2$, exposure cuts the amount of potential time one could live in half - that is, the failure time observed under constant exposure is half that we would observe under no exposure. To account for time varying exposures, we have to integrate the exposure function over the observed time.

This model allows us to calculate the potential failure time under no exposure using the data. For example, if $\psi = 0.2$ for an individual with an observed failure time $T = 2.2$ and exposure history $\bar{\boldsymbol{X}}_{\boldsymbol{k}} = (1,0,1)$, the survival time under the exposure history "never exposed" ($\bar{\boldsymbol{X}}_{\boldsymbol{k}} = (0,0,0)$) would be

$$T^{\bar{\mathbf{0}}} = e^{0.2*1} * (1-0)\text{years} + e^{0.2*0} * (2-1)\text{years} + e^{0.2*1} * (2.2-2)\text{years} = 2.47 \text{ years}$$

Estimation of the SNAFT model is done using g-estimation *Robins* (1989); *Robins and Tsiatis* (1992). A simple g-estimation algorithm would proceed as follows:

First, take a guess at ψ (called $\widetilde{\psi}$) and use equation 1.9 to generate a set of the potential failure times under no exposure at $\widetilde{\psi}$ called $\widetilde{T}^{\bar{\mathbf{0}}}$ as in

$$\widetilde{T}^{\bar{\mathbf{0}}} = \int_{k=0}^{T} \exp(\widetilde{\psi} X_k) d\,k.$$

Next, one uses an estimating equation to evaluate the association between $\widetilde{T}^{\bar{\mathbf{0}}}$ and

X_k. This could be a logistic model such as

$$logit[Pr(X_k = 1 | \boldsymbol{V}\bar{\boldsymbol{L}}_{\boldsymbol{k}}, \bar{\boldsymbol{X}}_{\boldsymbol{k-1}}, \widetilde{T}^{\bar{\mathbf{0}}}, T > k; \beta, \theta)] = \beta_{0k} + \boldsymbol{V}\beta_1 + \bar{\boldsymbol{L}}_{\boldsymbol{k}}\beta_2 + \bar{\boldsymbol{X}}_{\boldsymbol{k-1}}\beta_3 + \widetilde{T}^{\bar{\mathbf{0}}}\theta \quad (1.10)$$

in which the dependent variable is the exposure at time k (*Witteman et al.* (1998); *Hernán et al.* (2005)). The model includes a intercept β_{0k} that may be time varying. Because the outcome we would observe under no exposure can only depend on exposure if there is unmeasured confounding (i.e. $T^{\bar{\mathbf{0}}}$ is treated as a baseline variable determined prior to any exposure or covariates), if all confounders are accounted for in the right side of equation 1.10, $\widetilde{T}^{\bar{\mathbf{0}}}$ will equal $T^{\bar{\mathbf{0}}}$ when $\theta = 0$ (*Robins* (1989)). One can perform a grid search over a reasonable range of $\widetilde{\psi}$, and the $\widetilde{\psi}$ that yields a Z-statistic for θ equal to 0 (or very close to 0) is the estimate $\hat{\psi}$. G-estimation is the term given to this search.

To develop some intuition, one could consider $T^{\bar{\mathbf{0}}}$ to be the "residual" outcome once the net effects (the parameter of interest) of exposure are removed. It is the variation in the observed outcome that is not due to exposure. The ψ parameter corresponds to this net effect. Once we have removed the variation in the outcome not due to exposure, then the residual outcome should be independent with an increment of exposure (X_k). We might expect that the residual outcome would vary between individuals with different employment and exposure histories, so we test for this independence within strata of these variables.

The formulation of a SNAFT model described in this section assumes that all outcomes are observed. This model requires modification when when some outcomes are unobserved due to censoring. A description of the g-estimation of a SNAFT model in the presence of censoring can be found in *Witteman et al.* (1998), or in *Joffe et al.* (2012) (the latter of which is more technical).

The positivity assumption for g-estimation As discussed in Appendix C.3, the positivity assumption requires that we observe all levels of the exposure in all levels of the

covariates. Using the estimating equation 1.10 as an example, note that occupational studies are sensitive to this assumption. If exposure cannot occur off of work, then the probability of exposure when $L_k = 0$ (currently off work, a component of $\bar{\boldsymbol{L}}_{\boldsymbol{k}}$) will always be zero. In such a case, the coefficients β_1, β_2 and θ are not estimable in a saturated model.

G-estimation can overcome such violations of the positivity assumptions, provided we replace it with another assumption (*Joffe et al.* (2010)). An example of this assumption was given by *Chevrier et al.* (2012), who used a modified estimating equation of the form

$$logit\left[Pr(X_k = 1|\boldsymbol{V}, \bar{\boldsymbol{X}}_{\boldsymbol{k-1}}, \widetilde{T}^{\bar{\mathbf{0}}}, T > k, L_k = 1; \beta, \theta)\right] = \beta_{0k} + \boldsymbol{V}\beta_1 + \bar{\boldsymbol{X}}_{\boldsymbol{k-1}}\beta_2 + \widetilde{T}^{\bar{\mathbf{0}}}\theta$$

Where the estimating equation is only used on the person-years of active employment in the data ($L_k = 1$). Using occupational data, we can estimate all coefficients in this model because there is no longer a violation of positivity. This amounts to a relaxation of the assumption of no unmeasured confounding (see appendix C.3) so that it only applies to the person time at work. *Joffe et al.* (2010) refer to this as a selective ignorability assumption and offer several examples in which, even though the assumption of no unmeasured confounding does not hold for all data, information from a subset in which it does hold (e.g. subsets with better exposure measurement or no missing covariates) can be leveraged to estimate causal effects with less bias than would be possible using the full data.

An alternative to the approach of *Chevrier et al.* (2012) is to fit the model in equation 1.10 using an unsaturated model. In that case, empty strata will be smoothed over. However, in occupational data, the approach of *Chevrier et al.* (2012) is likely more defensible, since the smoothing done by the model would amount to allowing some individuals to be exposed off work. The smoothing approach is likely a good choice

when nonpositivity may happen because of sparse data, such as in analyses of data with long follow-up and many potential confounders.

The validity of this approach relies on specification of the structural nested model and the model proposed for the estimating equation. This is fewer models than are required by the g-formula, but more than a standard regression approach. Most examples of structural nested models in the literature are relatively simple models with dichotomous exposures (*Joffe et al.* (1998); *Hernán et al.* (2005); *Chevrier et al.* (2012); *Neophytou et al.* (2014)) and there are relatively few examples of fitting SNAFT models with quantitative exposures (*Joffe et al.* (2012); *Naimi et al.* (2014a)). Even in simulated data, SNAFT models may not perform well in reasonably sized data sets (*Young et al.* (2010)), and *Joffe et al.* (2012) noted difficulties with g-estimation when trying to estimate multiple SNAFT model parameters. However, these models may complement standard regression models, and they appear promising for wider use in occupational studies to control healthy worker survivor bias (*Joffe* (2012)).

1.5.2 Marginal structural models

Robins and colleagues also described marginal structural models, another class of models that can adequately control for time-varying-confounders that mediate the exposure-outcome association (*Robins* (1997, 1999); *Robins et al.* (2000)). A marginal structural model is a tool for analysis that estimates associations between exposure and an outcome, much like standard regression models. They are relatively simple in implementation and provide parameters with familiar interpretation. Thus they are an attractive tool for epidemiologists.

A trivial example of marginal structural model involves some exposure that is unconfounded (as treatment in a clinical trial) and all outcomes are observed (no censoring) (*Robins* (1997)). In that case, the marginal structural model is equivalent to a crude (i.e. marginal) model for the exposure. Observational studies are often subject

to confounding, however, so this trivial model is not usually of interest. Assuming that all confounders are measured, one can estimate a marginal structural model using inverse probability weighting (*Horvitz and Thompson* (1952)).

Returning to the notation used in previous sections (shown in table 1.2), these weights can be estimated using data via

$$\hat{W}_k = \prod_{j=0}^{k} Pr[X_j = x_j | \bar{\boldsymbol{L}}_{\boldsymbol{j}} = \bar{\boldsymbol{l}}_{\boldsymbol{j}}, \bar{X}_{j-1} = \bar{x}_{j-1}, \boldsymbol{V} = \boldsymbol{v}]^{-1}$$

Informally, inverse probability weighting relies on the principal that the same types of individuals are present in both exposed and unexposed groups, but in different proportions that can be weighted to a common standard. The weights are applied to data so that each individual is represented by $\hat{W}_k$ copies in the data. The weighting results in a "pseudo-population" in which confounders included in the weight model are balanced over the levels of exposure. An inverse probability weighted marginal structural model using weights $\hat{W}_k$ is just a crude model fit to the pseudo-population, rather than the observed population.

A non-trivial example of using inverse probability weighting is to fit a marginal structural Cox proportional hazards model. For a binary exposure, this model estimates the change in the relative hazard for a one unit increase in exposure. *Hernán et al.* (2000) interprets the coefficient from such a model as the "ratio of the mortality (hazard) rate at any time t had all subjects been continuously exposed... compared with the hazard rate at time t had all subjects remained unexposed" (p 563). This definition has not been previously extended to a cumulative exposure.

The positivity assumption for inverse probability weighting Note that if

$$1/\hat{W}_k = \prod_{j=0}^{k} Pr[X_j = x_j | \bar{\boldsymbol{L}}_{\boldsymbol{j}} = \bar{\boldsymbol{l}}_{\boldsymbol{j}}, \bar{X}_{j-1} = \bar{x}_{j-1}, \boldsymbol{V} = \boldsymbol{v}] = 0$$

for any individual - that is, the probability of being exposed (or unexposed) at time k is 0, then W_k is undefined (or infinite). It is also an indication that there are no individuals to which appropriate weights can be applied. This implies that a pseudo-population created with such weights contains strata of exposure in which the covariates are not balanced (thus not controlling bias due to confounding). The necessity that all individuals in a study must have a positive (non-zero) probability of exposure is known as the positivity assumption (*Cole and Hernán* (2008); *Westreich and Cole* (2010)) or experimental treatment assignment assumption (*Petersen et al.* (2012)).

Marginal structural models have been popular with epidemiologists since their introduction. However, their use in the occupational literature has only two examples. *Dumas et al.* (2013) estimated the effect of occupational exposure to asthmagens on asthma expression using marginal structural models in a clinical population. *Thygesen et al.* (2011) estimated the association between employment as an electrician and mortality using marginal structural models. *Robins et al.* (2000) anticipated this use of marginal structural models, but noted that the models should not be used in occupational data due to problems with nonpositivity. *Naimi et al.* (2011) confirmed that marginal structural models could not remove confounding in simulated occupational data. When employment status is considered a confounder of interest, it will result in nonpositivity when exposure cannot occur outside of employment. The recommendation of *Robins et al.* (2000) was based on the model of healthy worker survivor bias in which employment status is the confounder of interest. Both *Dumas et al.* (2013) and *Thygesen et al.* (2011) used inverse probability weighting in occupational scenarios in which nonpositivity may be of less concern, but they were likely limited in the extent to which they could control healthy worker survivor bias. Still, these examples suggest that there may be some questions in occupational epidemiology that can be answered well using marginal structural models.

CHAPTER II: METHODOLOGY

2.1 Overview

We propose two approaches from the causal inference literature for reducing potential bias due to healthy worker survivor bias in data from the Colorado Plateau uranium miners study. The first approach will utilize g-estimation of a structural nested accelerated failure time (SNAFT) model, while the second approach will involve a modified application of a marginal structural model (MSM) using imputed background radon exposure to avoid violations of the positivity assumption. Both SNAFT models and MSMs have been shown in other settings to appropriately adjust for time-varying confounding without inducing additional bias by conditioning on variables influenced by exposure to radon in a regression model (e.g. *Cole et al.* (2005, 2007); *Hernán et al.* (2005); *Chevrier et al.* (2012)).

2.2 Colorado Plateau uranium miners study

Mainly due to increased demand for uranium due to broader development of the nuclear program by the United States military, extensive mining and processing of uranium began in Colorado in the 1940s and 50s (*Ringholz* (2002)). Radiation exposure had previously been suspected of suspected of having a link to lung cancer in earlier decades in the Saxony region. However, radon concentrations in the Colorado Plateau mines were thought to be considerably lower than the mines of the Saxony region and did not warrant monitoring (*Schüttmann* (1993); *Jacobi* (1993)). No radon concentration measurements were taken until the 1950s, when the United States Public Health

Service (USPHS) began health surveys of miners that included sporadic measurement of radon concentrations in mine-shaft air. These initial studies led to the formation of a cohort of miners in the Colorado Plateau carried out by the USPHS and then by the National Institute for Occupational Safety and Health (NIOSH) (*NRC* (1988)). This study of the Colorado Plateau uranium miners cohort, which is the subject of the proposed analyses, has been subject to repeated follow-up (*Archer* (1962); *Wagoner et al.* (1964); *Lundin Jr et al.* (1969); *Archer et al.* (1976); *Whittemore and McMillan* (1983); *Hornung and Meinhardt* (1987); *Roscoe* (1997)) for continued exposure data and vital status, most recently by *Schubauer-Berigan et al.* (2009).

2.2.1 Cohort definition

The data comprise over 4,000 white and Native American male underground uranium miners in the Colorado Plateau uranium miners cohort recent follow-up for vital status. The cohort includes all miners from the original cohort (e.g. *Archer* (1962)) who responded to one or more surveys from the U.S. Public Health Service between 1 January, 1950 and 31 December, 1959 and who had at least one month of underground uranium mining employment before 1 January 1964 (*Archer et al.* (1976)). These data include miners with mining experience prior to date of entry into the study, which is defined as the date of first examination by the Public Health Service (*Lundin Jr et al.* (1969)). Each member of the cohort was followed until the earliest of death (N=2,428), loss-to-follow-up (N=14), or 31 December, 2005. The surveys from 1957 and 1960 covered approximately 90% of the miners in areas visited by the USPHS (*Roscoe* (1997)).

2.2.2 Data acquisition

The Colorado Plateau uranium miners cohort data are available from the National Institute of Occupational Safety and Health upon request.

2.2.3 Vital status and cause of death information

As of December 31, 2005, the 2,428 of the 3,358 white cohort members were known to have died over the course of 95,867 person-years. Follow-up through 1990 was performed using records from the Social Security Administration, the Internal Revenue Service, the National Death Index, and Health Care Financing Administration, and death certificates were coded by a qualified nosologist for cause of death information (*Roscoe et al.* (1989)). For miners who were not known to have died before 1 January, 1990, vital status and cause of death information was obtained using the National Death Index and the Social Security Administration's mortality file. Imperfect matches on date of birth, name, and Social Security number were manually reviewed. Fourteen miners who died before the advent of the National Death Index in 1979 were lost to follow up. Lung cancer was listed as the cause of death for 549 individuals, and cause of death could not be determined for 22 individuals (*Schubauer-Berigan et al.* (2009)).

2.2.4 Employment history data

Employment history data was included as part of the USPHS surveys of the 1950s, and subsequent surveys in the 1960s. Additionally, the USPHS supplemented the data from initial questionnaires through an annual census of all US uranium miners from 1954 through 1969 (*Archer et al.* (1976)).

2.2.5 Radon exposure data

Estimated monthly radon exposure in Working Levels for each miner were based on one of four methods: 1) actual measurement, 2) interpolation or extrapolation in time 3) geographic area estimation, or 4) pre-1950 (when air monitoring began) estimates were based on knowledge of ore, ventilation practices, and subsequent measurements (*Roscoe et al.* (1989)). Between 1950 and 1968, the USPHS helped to make nearly 43,000

measurements of radon-progeny concentration in 2,500 Colorado Plateau uranium mines. Nearly 50% of the area estimates for each mine were made by interpolation of measurements in the same year (*Whittemore and McMillan* (1983)). Because radon exposures are not limited to Uranium mines, the USPHS also recorded history of work in hard-rock mines, though no radon measurements were made in these mines and estimated radon exposures are generally low relative to exposure due to uranium mine work (*Whittemore and McMillan* (1983)). Monthly radon exposures through 1969 for each miner have been estimated based on the conducted measurements and estimates of area-level radon concentrations and work history data collected in the USPHS surveys. Radiation levels dropped sharply after 1967, when only 12% of the cohort remained working, so mining exposure histories are essentially complete for the cohort (*Archer et al.* (1976)). It should be noted that some cumulative exposures in this cohort are extremely high (> 5,000 WLM), which has been cited as a source of non-linearity of effect estimates in this cohort - as well as providing a challenge to external validity. *NRC* (1999) and *Lubin* (1994) restricted analyses to miner-years with fewer than 3,200 cumulative WLM.

2.2.6 Smoking data

Detailed smoking information was initially obtained at study entry and in subsequent follow up examinations (*Archer et al.* (1976)). Smoking start dates, stop dates, and estimates of smoking intensity in pack-years were obtained at each time of follow-up (*Whittemore and McMillan* (1983)). The most recent follow-up of the Colorado Plateau uranium miners data by *Schubauer-Berigan et al.* (2009) includes a final smoking history survey given in 1985 by NIOSH in which 2/3 of the miners, or next of kin proxies for deceased miners responded. Prior smoking was more prevalent among non-respondents than among respondents. Prior to 1985, the most recent survey to assess smoking status was in 1969 (*Hornung et al.* (1995)).

2.3 Background Radon

Estimates of background radon exposures can be imputed based on an assumption of a parametric distribution fitted to measured radon concentrations from the High-Radon Project, a project of the Lawrence Berkeley National Laboratory (*Price et al.* (2011)). Based on 1443 measurements made in Colorado between 1986 and 1987, the median alpha activity measured was 3.3 pCi/L (121 Bq/m^3). Figure 2.1 shows a histogram of the readings below 50 pCi/L (1780 Bq/m^3) with a fitted log-normal curve with geometric mean and standard deviation equal to those in the data (3.2, and 1.0). Individual background radon radiation will be imputed for each miner based on a draws from a log-normal distribution that approximately fits the data, and then averaged over the draws.

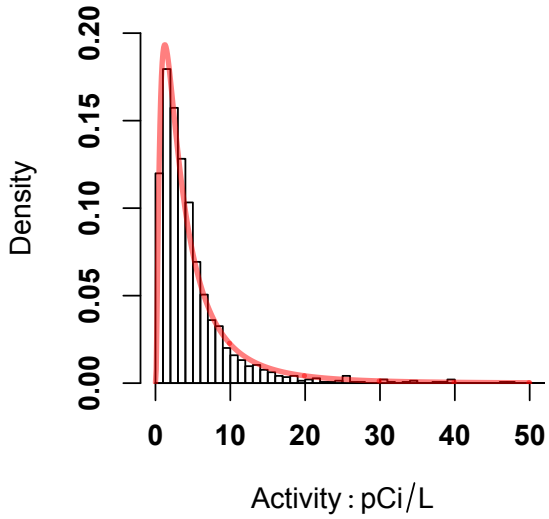

Figure 2.1: 1443 measured radon concentrations in 48 Colorado counties from 1986-1987 by the High-Radon Project, Lawrence Berkeley National Laboratory (*Price et al.* (2011))

2.4 Statistical Methods

The proposed analyses involve two estimation procedures for longitudinal data developed by *Robins* (1989, 1997), which are part of a larger class of statistical methods

for causal inference known as g-methods.

Both of these proposed methods are considered generally here and in more specific detail in §3.1 and 3.2.

2.4.1 G-estimation of structural nested accelerated failure time models

The simple SNAFT model model shown in § 1.5.1 can be generalized to continuous radon exposures, as observed in the Colorado Plateau uranium miners. We propose to first fit a basic SNAFT model

$$T^{\bar{0}} = \int_0^T \exp(\psi \bar{X}_{k-5}) d\,k$$

where $\bar{X}_{k-5}$ is the cumulative radon exposure in working level months up to time t with a lag of 5 years to allow for latency between lung cancer induction and mortality T. Because many time related factors mediate the radon-lung cancer association, We will also fit models of the allowing the time ratio to vary over time varying factors (time since exposure, age at exposure, smoking, exposure rate) of the relationship between radon and lung cancer. For example, to assess effects of time-since-exposure in correspondence with BEIR VI models, We will fit the model

$$T^{\bar{0}} = \int_0^T \exp(\psi_1 \bar{X}_{k-(5:14)} + \psi_2 \bar{X}_{k-(15:24)} + \psi_3 \bar{X}_{k-(25+)}) d\,k$$

Where $\bar{X}_{k-(a:b)}$ corresponds to the cumulative exposure accrued from a to b years prior to time k. We will estimate these models using the g-estimation procedure described in § 1.5.1. The structural model can also be used to explore lag functions, since the lag functions used or estimated under standard regression models (e.g.*Langholz et al.* (1999); *Hauptmann et al.* (2001)) will not necessarily hold for causal models. The lag can either be a fixed number, as expressed in the structural model here, or it can be

treated as a stochastic variable (*Richardson* (2009b); *Richardson et al.* (2011)).

Joffe et al. (2012) described a different set of structural models useful for allowing the effects of a continuous exposure to vary over time-varying factors. However, practical issues related to implementation of estimation algorithms limited exploration of effect measure modification in the analysis.

G-estimation of a SNAFT model in occupational data has previously been shown to be feasible by *Chevrier et al.* (2012). Additionally, the investigators showed several ways to generate other effect measures using ψ estimates from g-estimation. Because SNAFT models are well suited for overcoming the non-positivity that characterizes bias due to healthy worker survivor bias, We will estimate other effect measures for comparability, especially the hazard ratio as a method of evaluating the MSM described in § 2.4.2. An alternative approach is that taken by *Naimi et al.* (2014a), who compared SNAFT model parameters to parameters from a parametric accelerated failure time model.

2.4.1.1 Inference under measurement error of exposures and covariates

G-estimation can be used in any subset of the data for which conditional exchangeability holds, and *Joffe et al.* (2010) note that this can include a subset of the data in which the measurement error is less, thus reducing the impact of exposure misclassification. Based on the variation in measurement sources across time shown in figure C.2, the estimating equation could be limited to a subset of the data in which measurement error was least (e.g after 1950), but full inference for the data is still possible. This feature of g-estimation could also allow estimation in a subset with complete smoking information, thus extending to a number of possibilities regarding focused analyses on single modifiers or sensitivity analyses.

2.4.2 Marginal structural models using inverse probability weighting

The second proposed analysis of the Colorado Plateau uranium miners data involves a MSM of the form

$$\lambda_X(k) = \lambda_0(k)\exp(\beta \bar{X}_{k-5})$$

Where $\bar{X}_{k-5}$ is again the 5-year lagged radon concentration in WLM and T is the time to lung cancer, which We will estimate using a weighted Cox proportional hazards model with estimated weights $\hat{W}(k)$ as described in § 1.5.2, but using a continuous exposure.

$$\hat{W}_k = \prod_{j=0}^{k} \frac{f[X_j|\bar{X}_{j-1}, \boldsymbol{V}]}{f[X_j|\bar{\boldsymbol{L}}_{\boldsymbol{j}}, \bar{X}_{j-1}, \boldsymbol{V}]}$$

In principal, SNAFT models should more naturally allow for exploration of effect measure modification. However, as noted in the previous section, practical issues have limited the use of SNAFT models. The MSM may be the appropriate model to explore effect measure modification, in that case. While MSMs are noted to have shortcomings when addressing effect measure modification under certain circumstances (e.g. *Hernán et al.* (2001)), several authors have shown how to assess effect measure modification or to estimate joint effects using inverse-probability weighting (*Hernán et al.* (2001); *Petersen et al.* (2007); *Chiba et al.* (2009); *VanderWeele and Vansteelandt* (2011)).

As in the SNAFT model, the MSM is potentially useful for estimating a latency function, particularly when considering a weighted exposure history where weights are proportional to the expected contribution to disease at a given time. These latency model weights work naturally with the inverse probability weights (*Langholz et al.* (1999); *Richardson* (2009b)). Further, inverse probability weights have not been applied previously to a linear excess relative risk model, and the model is relatively straightforward to estimate once weights are estimated. This model may provide a useful comparison to previous analyses.

Because inverse probability weighting is sensitive to violations of the positivity assumption, non-positivity due to the lack of off-work exposures must be addressed in this analysis. We propose estimating the effect of total radon exposure (occupational + residential) using existing background radon estimates from *Price et al.* (2011). As an example, residential monitoring data for the state of Colorado (which holds part of the Colorado Plateau) is shown in figure 2.1. By combining residential with occupational exposure, the conditional probability density of exposure $f[X_k|\cdot]$ will be > 0 for all participants. These data are approximately log-normally distributed, and adding them to the Colorado Plateau uranium miners exposure estimates will likely create difficulties estimating the exposure density, given that the true density will be the sum of two approximately log-normally distributed variables.

CHAPTER III: RESULTS

3.1 Structural nested accelerated failure time models to control healthy worker survivor bias (AIM 1)

3.1.1 Abstract

Cohort mortality studies of underground miners have been used to estimate the number of lung cancer deaths attributable to radon. In the US, for example, it has been estimated approximately 20,000 lung cancer deaths per year are attributable to radon. However, the findings of prior analyses of radon-lung cancer associations among underground miners may be subject to healthy worker survivor bias, a type of time-varying confounding by employment status. This bias is problematic for standard regression models because radon exposure may influence employment status. We re-examined radon-lung cancer mortality associations in the Colorado Plateau uranium miners cohort study, one of the major cohorts included in prior pooled analyses of associations between radon and lung cancer among miners. Accelerated failure time models were fit to estimate the time ratio (the relative decrease in the median survival time) per 100 working level months (WLM - radon exposure averaging 130,000 MeV of potential alpha energy per liter, per working month). We adjusted for healthy-worker survivor bias using G-estimation of structural nested models. After controlling for this bias, the time ratio (95% confidence intervals) per 100 WLM was 1.168 (1.152, 1.174). In an unadjusted model the estimate was 1.102 (1.099, 1.112), a 39% decrease. We observed a similar decrease for all-cause mortality. We estimate that among 617 miners who died from lung cancer, 6,071 person-years of life were lost due to occupational radon exposure. Our analysis suggests a healthy worker survivor bias in standard analyses of the

miner data, warranting re-examination of current estimates of radon's carcinogenic effects.

3.1.2 Background

Radon is a ubiquitous gas that concentrates in indoor air and is a leading cause of lung cancer in the United States. Reliable estimates of the burden of lung cancer attributable to residential radon exposure is of considerable interest, given the high potential costs of compliance with the current EPA action level for radon, on the one hand, and the potential public health impacts of reducing that action level (as suggested by the Presidents Cancer Panel) on the other hand. (*Krewski et al.* (2005); *Darby et al.* (2005); *USEPA* (2003)). Radon exposure is protracted and may have persistent effects, so researchers frequently model radon-lung cancer associations using a cumulative metric of radon exposure. Among the most influential estimates of radon-lung cancer associations, for regulators and others involved in risk assessments, have been results of occupational cohort mortality studies of underground miners (*USEPA* (2003); *NRC* (1988); *World Health Organization* (2009); *ICRP* (2010); *NRC* (1999)). Typically, occupational studies are better suited than residential studies for estimating precise dose-response parameters (*NRC* (1999)); however, occupational studies are subject to unique biases that can reduce the applicability of their results to other settings.

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 (*Arrighi and Hertz-Picciotto* (1994)). The disease rates of employed and unemployed are generally not comparable, even among those with identical cumulative exposure. Issues of comparability in the radon literature have typically been addressed using multivariable regression models that include population predictors

of cancer outcomes, such as age and birth cohort (*NRC* (1999); *ICRP* (2010); *USEPA* (2003); *Hornung et al.* (1998, 1995)).

Healthy worker survivor bias can be conceptualized as a form of confounding by employment status (*Robins* (1986)). The health related factors that underlie the ability to stay at work may also influence the outcome under study. Multivariable regression methods used in previous studies can control this confounding in some cases. However, regression methods cannot completely control this bias when exposure in the past affects subsequent employment (Figure 3.1) (*Pearce* (1992)). The potential for this bias has not been evaluated in miner studies, though methodologic advances have made such an evaluation possible (*Robins and Tsiatis* (1992)). We estimate dose-response parameters between radon and lung cancer and all-cause mortality using methods developed to control confounding when the confounders may be affected by past exposure.

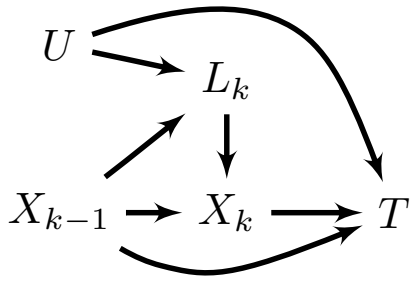

Figure 3.1: Causal diagram showing hypothesized relationships that necessitate use of special methods to control healthy worker survivor bias in the Colorado Plateau Uranium Miners data. X_k is radon exposure and L_k is active employment status in month k.

3.1.3 Materials and methods

3.1.3.1 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 (*Holaday et al.* (1957)).

Follow-up for mortality was assessed through 12/31/2005 (*Roscoe* (1997); *Schubauer-Berigan et al.* (2009)). Cause of death information was obtained 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 MeV of potential alpha energy per liter of air] were derived from raw data files (*Schubauer-Berigan et al.* (2009)). These exposure data were originally derived from a job-exposure matrix for the years 1950-1969 using area measurements and extrapolations from nearby mine-shafts, mines, or regional averages. In 1950, 25% of the exposures were derived from expert opinion. Estimated radon exposure due to previous work in hard-rock (i.e. non-uranium) mines was included as a covariate. Three miners were excluded who had lifetime cumulative exposures greater than 10,000 WLMs and employment time that leads to an average exposure of 1 working level over 170 hours).

Individual information on smoking histories was obtained from surveys conducted in 1985. 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 minimum of death, loss to follow-up, or 12/31/2005.

3.1.3.2 Statistical Methods

We estimate the change in the expected age at death due to an increment of cumulative radon exposure under a linear dose-response assumption by fitting an accelerated failure time model. This quantity is expressed as the time-ratio (TR the relative decrease in the median survival time for an additional unit of exposure) and is reported along with associated 95% confidence intervals (CI) for a 100 WLM increase in cumulative radon exposure. Analyses examined death due to lung cancer and all-cause mortality. 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 (as we have estimated it, in which a value > 1 indicates harmful exposure) and hazard ratio will be identical, though this equivalence does not hold for other distributions (*Kalbfleisch and Prentice* (2002)). 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 estimated the TR using a structural nested accelerated failure time model (SNAFT model) fit by g-estimation (*Robins* (1989)). Here we provide a basic explanation of the SNAFT model in a study with no censoring; we fully describe our approach with the miner data in the online appendix.

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. Therefore, we considered exposure estimates and employment duration before study entry to be time-fixed covariates. 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:

$$T^{\bar{\mathbf{0}}} = m + \int_{m}^{T} (1 + \psi \bar{X}_{k-60}) d\,k \tag{3.1}$$

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, ψ is the parameter of interest and $T^{\bar{\mathbf{0}}}$ is the expected survival time, given no radon exposure during follow-up. Time is denoted by k.

Consistent with much of the prior radon literature, in which the excess relative rate (rate ratio - 1) is modeled on a linear scale (*Lubin* (1988)), the parameter ψ increases linearly with exposure and is defined as the excess relative time (TR-1). This approach contrasts with previous uses of SNAFT models, which are typically log-linear (e.g. *Hernán et al.* (2005)). As a technical note, our model places no bounds on ψ (unlike the constraint the must be placed upon the parameter describing the radon dose effect in a linear excess relative rate model), so our model potentially allows for a negative time ratio in the case of beneficial exposures.

Usually, accelerated failure time models require specification of the baseline survival distribution for the disease of interest. In SNAFT models this distribution can be left unspecified. In these models, the baseline time, $T^{\bar{\mathbf{0}}}$, can be interpreted as a potential outcome. The outcome represents the time of death we would observe, had we intervened to prevent exposure at work (for example, by a ventilation system that completely replaced mine air with air containing no radon). This interpretation allows one to easily calculate the years of life lost (among cases) due to exposure as $T - T^{\bar{\mathbf{0}}}$, which we use to supplement the TR as an estimate of the impact of radon exposure. We calculated the years of life lost due to exposure for all-cause mortality and lung cancer mortality.

$T^{\bar{\mathbf{0}}}$ is treated as an individual level variable that can be deterministically derived

from the model shown in (1) positing a value for ψ and using the observed quantities: age at death, cumulative exposure, and age at entry. The best estimate of the parameter ψ is obtained by iteratively searching for the value of ψ at which the $T^{\bar{0}}$ is independent of monthly radon exposure X_k, conditional on covariates.

Testing the conditional independence of $T^{\bar{0}}$ and X_k can be done by including the potential outcome as a covariate in a model that predicts monthly exposures (the exposure model), conditional on prior covariates. The coefficient for $T^{\bar{0}}$ in the exposure model can be used to test this conditional independence. At the best estimate of ψ, the variation of radon exposure within groups of similar individuals in any given month should not be associated with $T^{\bar{0}}$. A point estimate and associated 95% confidence interval for ψ was obtained using a line-search algorithm that iteratively fit the SNAFT over a range of values for ψ (*Hernán et al.* (2005)). Under our model, a TR greater than 1 indicates a harmful exposure. Technical details of the exposure model are shown in the appendix.

3.1.3.3 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. Current employment status was controlled for by restricting the exposure model to periods of active employment (i.e. $L_k = 1$) and we adjust for exposure and employment history $(\bar{X}_k, \bar{L}_k)$ by including a terms for the history variables described in Appendix Table A.1 up to, and including time k. 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 the covariates fixed at study entry.

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 $100\% * (\psi_{adjusted} - \psi_{unadjusted})/\psi_{adjusted}$. A negative value was interpreted as evidence that the radon effect is underestimated due to healthy worker survivor bias.

In order to describe variation in radon-lung cancer association with time since exposure, similar to previous analyses we also estimated the TR for windows of exposure from the preferred model of the Committee on the Biological Effects of Ionization Radiation (*NRC* (1999)). This analysis involves a model of the form:

$$T^{\tilde{0}} = m + \int_{m}^{T} (1 + \psi_1 \bar{X}_{k1} + \psi_2 \bar{X}_{k2} + \psi_3 \bar{X}_{k3}) d\,k \tag{3.2}$$

Where $\bar{X}_{k1}, \bar{X}_{k2}$, and $\bar{X}_{k3}$ correspond to the exposure accrued (since follow up began) between 5-14 years, 15-24, and 25+ years prior.

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 (*Applebaum et al.* (2011)), 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.

3.1.4 Results

3.1.4.1 Demographics and exposure distribution

Our cohort comprised 769 non-white miners and 3355 white miners who were followed for 27,343 and 107,626 person-years (Table 3.1). No cause of death could be determined for 22 miners and 14 were lost to follow-up before 1979. A majority (70% non-

white, 75% white) of the miners died before 12/31/2005, and 8.1% of non-whites and 16.5% of whites died of lung cancer (a difference previously attributed to differences in smoking patterns) (*Roscoe et al.* (1995)). Non-white miners were followed-up for longer (median 38.6 years versus 34.1 years for white miners) and worked longer during follow-up than white miners (median 10.8 years versus 7.4 years for white miners), despite similar employment time before follow-up (2.5 and 2.4 years for non-whites and whites). 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 study entry 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 (2.4 X 100 WLM vs. 0.76 X 100 WLM for non-whites and 1.9 X 100 WLM vs. 1.2 X 100 WLM for whites). Monthly exposure distributions were highly right skewed (Figure 3.2).

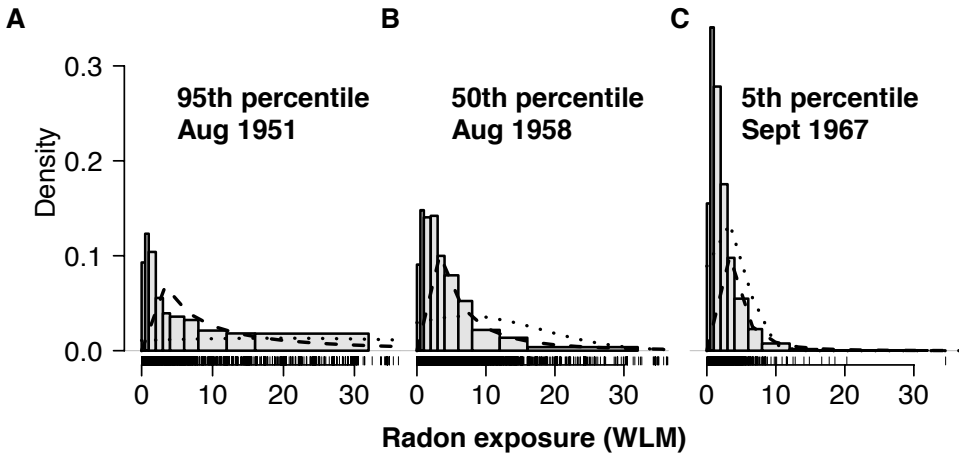

Figure 3.2: Illustrative monthly exposure distributions for white males born from 1920-1929 up to 35 WLM. Months selected represent the 5th, 50th, and 95th percentile for the mean monthly exposure from 1950-1969. Lines of fit are for visualization only and do not represent actual fit of exposure models. Lines below histogram represent actual monthly exposures for individual miners.

Table 3.1: Demographic and exposure distributions in the miner data

		White N=3355 107,626 PY	**Other Race** N=769 27,343 PY
		N(%)	
Vital status*	Alive	781 (23.3)	208 (27.0)
	Deceased - unknown cause	51 (1.5)	20 (2.6)
	Deceased- known cause	2514 (74.9)	535 (69.6)
	Deceased - lung cancer	554 (16.5)	63 (8.1)
	Lost to follow-up	9 (0.0)	6 (0.1)
Birth cohort	<1900	171 (5.1)	21 (2.7)
	1900-1909	460 (13.7)	76 (9.8)
	1910-1919	857 (25.5)	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)
Age at first uranium mining exposure	Never exposed	5 (0.1)	3 (0.4)
	<35	2113 (62.9)	565 (73.5)
	35-54	1132 (33.7)	188 (24.5)
	55+	105 (3.1)	13 (1.7)
		Median (IQR)	
Years of follow-up	Cases**	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**	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 baseline	Cases**	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**	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 exposure during follow-up (100 WLM)	Cases**	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 baseline (X100 WLM)	Cases**	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	0.30 (0.00, 17.8)	0.00 (0.00, 0.00)

* Vital status as of 31 December, 2005

**Cases = individuals with underlying cause of death listed as lung cancer

3.1.4.2 Dose-response analyses

Using a model for all-cause mortality under a 5-year cumulative radon exposure lag, we estimated the adjusted TR (95% CI) per 100 WLM was 1.054 (1.041, 1.068), while the unadjusted TR was 1.012 (1.013, 1.015) a decrease of 74% (Table 3.2). 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).

Table 3.2: Time ratios and 95% confidence intervals per 100 working level months, 5 year lag. Male uranium miners, Colorado Plateau, USA 1950-2005.

		Lung Cancer				All Causes			
Model*		TR	95% CI		% diff†	TR	95% CI		% diff†
Adjusted*	1	1.167	(1.152,	1.174)	ref	1.054	(1.041,	1.068)	ref
Unadjusted**	3	1.102	(1.099,	1.112)	-39	1.014	(1.013,	1.015)	-74

* 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, current employment status, and cumulative time at work during follow-up.
** Adjusted for: baseline exposure from uranium mining, race, prior mining exposure, birth cohort, date of hire
† Percent difference in ϕ from fully adjusted model

For lung cancer, the adjusted TR (95% CI) per 100 WLM was 1.168 (1.152, 1.174) while that for the unadjusted model was 1.102 (1.099, 1.112) a decrease of 39% (Table 3.2). Based on our adjusted model, we estimate that, among 617 miners who died from lung cancer during follow-up, occupational radon exposure after enrollment was associated with 6,071 person-years of life lost due to premature lung cancer death (not shown).

To evaluate variation in the dose-response for lung cancer with time-since-exposure we fitted a model with exposure time-windows. The estimated TR (95% CI) for lung

cancer per 100 WLM was 1.188 (1.116, 1.230) for exposures 5 to 15 years prior, 1.128 (1.050, 1.294) for exposures, 15-25 years priors, and 1.022 (0.950, 1.198) for exposures 25+ years prior (Table 3.3).

Table 3.3: Time ratios and 95% confidence intervals for windows of exposure from adjusted model for lung cancer

Exposure window*	TR**	95% CI	
5-15 years	1.188	(1.116,	1.230)
15-25 years	1.128	(1.050,	1.294)
25+ years	1.022	(0.950,	1.198)

* Cumulative exposure within the exposure window
**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.

To evaluate the sensitivity of our results to including prevalent hires in the cohort analysis we repeated analyses after excluding people who had long durations of employment prior to start of follow-up. To prevent model instability from sparse strata in this analysis, we reduced the number of birth cohorts to 3 (from 8 in the main analysis). This reduced the TR relative to our main analysis. The adjusted TR decreased slightly after excluding workers with substantial employment before study entry (Table 3.4).

G-methods are necessary 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 (*Naimi et al.* (2013)), 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 all covariates including employment history. The hazard for terminating employment was lower in workers with cumulative radon exposure above the median (1.2 X 100 WLM) vs. those with cu-

Table 3.4: Sensitivity analysis for the change in the TR of the radon/lung cancer association by excluding short to long term prevalent hires in the study cohort.

Max. employment prior to enrollment*	TR	95%CI	
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
Full cohort**	1.095	1.087	1.117

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

**Birth cohort was represented by 3 indicator variables (versus 8 in the main analyses), resulting in different TRs between these analyses and analyses reported in Table 3.2

mulative exposure less than the median (referent); HR (95% CI) = 0.90 (0.84, 0.98), not shown. Second, we fit a model for the comparing the hazard of death between not employed as a uranium miner and employed (referent) person time, adjusted for covariates including cumulative exposure with a lag of 2 years; HR (95% CI) = 3.3 (2.4, 4.3). Thus, parametric models adjusting for employment history will be biased and SNAFT models are needed to appropriately adjust for time-varying confounding by employment status.

3.1.5 Discussion

Healthy-worker survivor bias can occur in occupational studies when exposure accrues over time and workers with better prognoses are less likely to leave employment. Compared to models in which we adjusted for healthy worker survivor bias, we observed a 39% decrease in the TR for lung cancer and a 74% decrease in the TR for all-cause mortality in unadjusted models. Our analysis supports previous speculation

that there may be healthy worker survivor bias in the Colorado Plateau uranium miner data (*Stayner et al.* (2003)). Further, our analysis suggests this bias results in an underestimate of the dose-response between radon and both lung cancer and all-cause mortality.

Previous authors have conceptualized healthy worker survivor bias as a result of time-varying confounding by employment status (*Robins* (1986, 1987b); *Pearce* (1992)). Controlling this confounding by including employment status in a regression model for the RR (or TR) is potentially biased because employment status may be affected by prior exposure. Any model with cumulative exposure stratified by employment status will be stratifying the RR (or TR) by an effect of the exposure, leading to bias of the dose-response (*Rosenbaum* (1984); *Weinberg* (1993); *Robins and Wasserman* (1997)). We observed that prior radon exposure is associated with leaving work in the miner data, so multivariable regression models will be subject to this bias. SNAFT models can adequately control healthy worker survivor bias in this scenario because the models achieve confounder control without stratification (*Robins* (1989)).

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 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 model may be biased downward due to confounding by employment status, and those from the exposure-age-duration model may be biased after stratifying by a proxy for employment history.

We assessed the magnitude of healthy worker survivor bias by comparing an adjusted model to an unadjusted model. Other authors have assessed this bias by com-

paring SNAFT models with Cox regression models (*Chevrier et al.* (2012)). or parametric accelerated failure time models (*Naimi et al.* (2014a)). In contrast, our approach uses a straightforward comparison of two SNAFT models. However, our analysis remains sensitive to misspecification of the exposure model. Part of the difference between the adjusted and unadjusted SNAFT models could be due to misspecification of one of the exposure models. To reduce issues of model misspecification, most previous examples have used simpler models than our own, by fitting models for binary exposures (*Robins et al.* (1992); *Witteman et al.* (1998); *Joffe et al.* (1998); *Keiding et al.* (1999); *Korhonen et al.* (1999); *Cole and Chu* (2005); *Hernán et al.* (2005)). or exposure quantiles (*Naimi et al.* (2014a)). In contrast, we report SNAFT models under a zero-inflated, log-linear model for exposure, which allows for a skewed exposure distribution and inclusion of unexposed individuals (*Li et al.* (2011)). In our online appendix, we also fit SNAFT models under alternative exposure models. While results are somewhat sensitive to the choice model, our preferred exposure model fit the data best (based on Aikiake's Information Criteria) among the candidate models that we evaluated.

We estimate the effects only of exposure that occurs following the initial health interview. Because radon exposure itself may cause attrition from the workforce, the miners in our dataset were subject to a pre-enrollment selection process that may introduce bias. 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. To address these possible differences, we adjusted for pre-enrollment exposure and employment history as baseline 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 (*Suissa* (2008)). Estimating a dose-response using only post-enrollment exposures is not conventional in the miner literature, but it is not clear from published reports whether studies of major miner cohorts are subject to bias from immortal person-time. To illustrate the po-

tential 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 TR-1 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 (*Applebaum et al.* (2007)). As shown in Table 3.4, the apparent magnitude of healthy worker survivor bias decreases after excluding workers with long periods of employment before follow-up, suggesting that this bias is stronger among the prevalent hires. The stronger bias among prevalent hires 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 exclusions. 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.

Smoking is a strong risk factor for both of our outcomes of interest. In our data, smoking data were captured via a 1985 survey of living miners or proxy interviewees (*Schubauer-Berigan et al.* (2009)). Consequently, we do not have adequate information to evaluate the role of smoking as a time-varying confounder. Previous analyses have suggested that smoking may modify the effect of the radon-lung cancer association (*Tomásek* (2011)), but is not a source of strong time-fixed (*Leuraud et al.* (2011)). or time-varying confounding (*Richardson et al.* (2012)). In our context, smoking may affect both employment status and the outcomes under study, a point previously raised by Pearce and colleagues (*Pearce et al.* (2007)). SNAFT models can adequately control this bias by adjustment for employment history, if we assume that smoking is not asso-

ciated 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 (*Richardson et al.* (2012); *Lubin et al.* (1995b)). As we have shown, SNAFT models are well suited to address questions regarding time-varying covariates. Unfortunately, the estimation algorithm did not converge in a SNAFT model to estimate modification of the TR by exposure concentration, so we were unable to assess the TR over levels of exposure concentration (not shown). This problem echoes previous difficulties raised with addressing modification in SNAFT models raised by Joffe et al, and may be a shortcoming of using SNAFT models in practice (*Joffe et al.* (2012)). The SNAFT model yields a TR that is marginalized with respect to the distribution of confounders in the study population, so TR heterogeneity by exposure concentration would not explain the different results between our adjusted and unadjusted SNAFT models. Our models fit using time windows of exposure agreed qualitatively with previous analyses (*Langholz et al.* (1999); *Tomásek* (2012)), suggesting that correction for healthy worker survivor bias does not alter the apparent decrease in the impact of cumulative exposure on mortality with time since exposure.

3.1.6 Conclusion

SNAFT models are useful for estimating a dose-response for cumulative exposures in occupational studies. While G-estimation provides unique challenges in estimation and requires more computational resources and analyst time than most regres-

sion analyses, it does not suffer from the pitfalls that may prevent regression models from adequately controlling healthy worker survivor bias. 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 (*Armstrong* (1998)), and biases the dose response (*Stram et al.* (1999)), co-exposure to other lung carcinogens such as arsenic (*Arrighi and Hertz-Picciotto* (1996)), diesel exhaust, or silica (*Bergdahl et al.* (2010)). 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. Our analysis demonstrates that current risk estimates for radon based on occupational studies may not accurately reflect the net effects of radon exposure. We show evidence that healthy worker survivor bias has resulted in underestimates in prior studies, and improved handling of employment history as a confounder is a necessary step in reducing this bias.

3.2 Marginal structural models in occupational cohort studies (AIM 2)

3.2.1 Abstract

Occupational epidemiologists are often concerned with estimating the impact on health of exposures within the workplace. It is common practice to form aggregate measures of exposure over time, such as cumulative exposure. When there are factors, such as employment status, that strongly influence exposure and are associated with the outcome of interest, confounding may result. If exposure influences these factors, then regression models will be biased. Marginal structural models, usually fit by inverse probability weighting, were developed to address confounding when the confounders may be influenced by prior exposure. However, Robins and colleagues (2000) noted that a technical issue with inverse probability weighting negates the use of marginal structural models in occupational studies. We explore these issues and recommend situations in which marginal structural models may be of use to occupational epidemiologists. We ground our discussion in a small synthetic example and include an example using the Colorado Plateau uranium miners cohort. In the miners cohort, we show that efforts to avoid technical issues with inverse probability weights may induce bias that negates the control of confounding. We suggest a limited list of occupational settings in which inverse probability weighting can be used. Further, we suggest that marginal structural models fit using the parametric g-formula may be a way to implement this useful statistical tool to reduce bias in occupational studies.

3.2.2 Introduction

Occupational epidemiologists are often concerned with estimating the impact on health of exposures within the workplace. When these exposures have persistent health effects, it is common practice to form aggregate measures of exposure over time, such as cumulative exposure. This approach is attractive in that effect measures of cumula-

tive exposures are easily interpretable and easy to communicate. The simplicity of this approach comes with costs, however. Cumulative exposures in occupational settings are determined partly by employment duration. As a consequence, dose-response analyses of cumulative exposures in occupational cohorts are subject to bias when employment duration is affected by health. Healthier workers will tend to remain employed longer than less-healthy workers and can be exposed for longer durations and to higher cumulative amounts of workplace hazards (*Arrighi and Hertz-Picciotto* (1994)).

The underlying factors that influence both employment and the health outcomes of interest are often not recorded in occupational data, even if some are known. In lieu of such measures, a common approach is to treat this bias as confounding by employment status (*Gilbert* (1982)). Bias of cumulative exposure effect estimates from time-varying confounding by employment status is often referred to as healthy worker survivor bias (*Buckley et al.* (2014)).

Reducing healthy worker survivor bias is not straightforward when occupational exposures may be associated with subsequently leaving work (*Pearce* (1992)). Marginal structural models are a set of analytic tools to address this form of time-varying confounding in which prior exposure may affect the confounder at subsequent time points. Thus, these models appear to be a promising method to this bias. However, in their seminal paper on marginal structural models, Robins, Hern{'an and Brumbeck (2000) note that these models should not be used in occupational cohort studies due to technical issues with the method commonly used to fit the models, inverse probability weighting (*Robins et al.* (2000)). In the current manuscript, we study the use of marginal structural models for occupational studies of cumulative exposure-disease associations. We first revisit the recommendation made by Robins, Hernán and Brumbeck; next we refine its scope; and, finally, we give illustrative examples in which marginal structural models will, and will not, succeed in reducing bias in analyses of occupational studies.

3.2.3 Methods

We consider the setting of an occupational cohort followed over time and assessed for a continuous health outcome at the end of follow up (e.g. systolic blood pressure). To simplify our discussion, we focus on a hypothetical cohort followed for two years. Within this cohort we measure exposure to an occupational agent, X, employment status, L, and disease status, Y. Letting subscripts index the year of the study, all workers enter the study employed ($L_1 = 1$) and on the first annual visit a worker may be exposed ($X_1 = 1$) or unexposed ($X_1 = 0$). For the second visit, our data include exposure (X_2) and employment status [left employment just prior to year 2 ($L_2 = 0$) or remained employed ($L_2 = 1$)]. An employee who left work at the start of year 2 was considered to be unexposed in that year.

We are interested in the net effect of cumulative exposure ($X_k = X_1 + X_2$) on the outcome (Y) at the end of year 2. Without sacrificing generality, we focus on estimating the difference in the mean of the outcome per unit increase in cumulative exposure (hereafter, mean difference).

In Table 3.5, we show data on cumulative exposure, employment status, and the outcome in this hypothetical study. A crude linear regression model yielded a mean difference of 8, while a model adjusted for employment status yielded a mean difference of 11 (Table 3.5). However based on the relationships underlying our data shown in Figure 3.3, both of these estimates may be biased.

The crude estimate is subject to confounding because L_2 (the confounder) affects exposure in year 2 and the outcome of interest. The estimate adjusted for L_2 may be biased because L_2 is also a causal intermediate for exposure in year 1 (*Rosenbaum* (1984)). This is the underlying problem for some cohort studies in which healthy worker survivor bias is of concern.

Table 3.5: Mean outcome ($E[Y]$) by cumulative exposure $\left(\sum X_k\right)$and employment status (L_2) in a hypothetical occupational cohort

$\sum {X_k}^*$	L_2	N	$E[Y]$
0	0	200	35
0	1	100	11
1	0	400	43
1	1	200	28
2	1	60	37

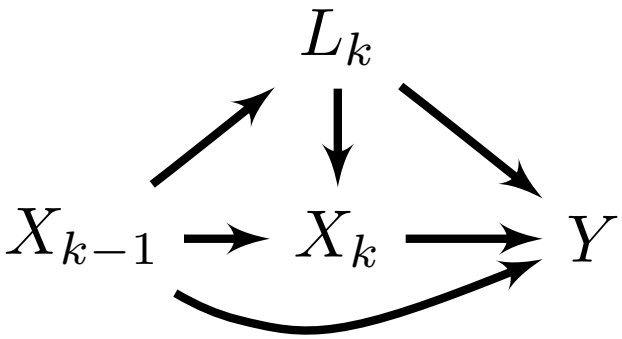

Figure 3.3: Diagram showing causal structure underlying the hypothetical example shown in tables 1 and 2. This diagram reflects hypothesized mechanisms underlying healthy worker survivor bias (*Buckley et al.* (2014)).

3.2.4 Marginal structural models

Marginal structural models can be used to estimate the effect of an exposure of primary interest when there is a confounder that is affected by prior exposure, as in Figure 3.3. Marginal structural models are typically fit using inverse probability of exposure weights. These models yield an effect measure that will be unconfounded by the factors defining these weights, including by time-varying confounders. Informally, the weights (W^x) are defined as the inverse probability of receiving ones own exposure history. For example, for a worker with $X_1 = 1, X_2 = 0$, W^x is one over the probability of having had that particular set of exposures, given the set of possible exposure histories and the observed confounders. Weights are typically derived through a model for the probability of exposure, conditional on previous values of exposure and confounders (*Cole and Hernán* (2008)). Making W^x copies of each observation yields a pseudo-population in which confounders are balanced over levels of exposure. This is analogous a clinical trial in which treatment is randomized and one expects covariate balance across treatment groups. An unconfounded estimate of association can be obtained by fitting a crude (or baseline adjusted) model to the pseudo-population data rather than the observed data. A marginal structural model effect estimate an interpretation as a standardized effect estimate (*Hernán and Robins* (2006)).

3.2.5 Marginal structural models in occupational studies

Returning to our example, we wish to estimate the net effect of cumulative exposure ($X_k = X_1 + X_2$) on the outcome using the data in Table 3.5. In Table 3.6, we display the same data in a slightly different manner. We have disaggregated exposure to show the calculation of the inverse probability weights needed for a marginal structural model.

We define our inverse probability weights as $W^x = W_1 W_2$ where

$$W_2 = Pr(X_2 = x_2 | L_2 = l_2, X_1 = x_1)^{-1}$$

Table 3.6: Mean outcome ($E[Y]$), stabilized inverse probability weights (SW^x) and pseudo-population cell sample size by strata of exposure (X_1, X_2) and employment status (L_2) in a hypothetical occupational cohort

X_1	L_2	X_2	$E[Y]$	N	SW^{x*}	$pseudo-N^*$
0	0	0	35	200	0.63	125
0	0	1	?	0	**	**
0	1	0	11	100	1.75	175
0	1	1	29	180	0.58	105
1	0	0	43	400	0.88	350
1	0	1	?	0	**	**
1	1	0	19	20	3.50	5
1	1	1	37	60	0.17	75
Overall†				960	0.87	835

* Pseudo-population cell sample size ($SW^x N$)
** Division by zero yields undefined weights and pseudo-population cell sample size
† $SUM(N)$, $SUM(pseudo-N)$, $MEAN(SW^x)$

and W_1 is a constant that can be ignored because exposure is randomized in the first year this will not hold in general but it simplifies our example. Lowercase letters are used to denote realizations (e.g. 1 or 0) of each variable. In table 2, we report stabilized weights SW^x, where $SW^x = SW_1 SW_2$ and $SW_2 = W_2 Pr(X_2 = x_2 X_1 = x_1)$ SW_1 is equal to one in our example, so $SW^x = SW_2$. The weight is undefined in all strata in which $L_2 = 0$ and $X_2 = 1$ (off-work, exposed). This occurs because no unemployed individuals are exposed, yielding a zero in the denominator.

For each row j in table 3.6, the pseudo-population size, pseudo-N, is defined as $SW_j^x N_j$. The mean of the stabilized weights should equal 1.0, implying that the total pseudo-population size should equal the size of the cohort, N. When the mean of the stabilized weights does not equal 1.0, it can serve as an informal diagnostic of problems with estimation of the weights.

A marginal structural linear model fit using these weights yields a mean difference

of 11.3 (Table 3.8). However, we have strong indications that this estimate is biased – the mean stabilized weight is 0.87, while it should equal 1.0.

3.2.6 Nonpositivity

Robins and colleagues (2000) noted that marginal structural models should not be used in occupational cohort studies (p 557) because of a common problem with estimating inverse probability weights, sometimes termed nonpositivity (*Robins et al.* (2000)). The condition that exposure must take on all possible values in all strata of confounders is known as positivity. Violations of this condition are referred to as nonpositivity (*Westreich and Cole* (2010)). This is exactly the problem we are experiencing in the hypothetical cohort.

As a heuristic example, if inverse probability weighting works in our cohort, employment status in year 2 should be balanced over levels of exposure in the pseudo-population. When a binary exposure can only take one value in some strata of confounders, weighting does not produce this balance and thus does not remove confounding (*Cole and Hernán* (2008)). In our example, software will yield a pseudo-N of 0 in all strata of exposed, unemployed. In that case, the proportion of workers employed in year two in the pseudo-population will be (175+70)/(175+70+350+125)=**0.34** for $X_2 = 0$ but it will be (105+10)/(105+10)=**1.0** for $X_2 = 1$. Because 0.34 does not equal 1.0, the weights have failed to balance the covariate over levels of exposure and confounding remains.

Nonpositivity can occur in any study due to sparse data within strata of confounders. In occupational studies, nonpositivity can occur more systematically. Those that leave employment are often assumed to be unexposed. This leads to nonpositivity when one includes employment status as a confounder.

3.2.7 Appropriate uses of marginal structural models in occupational studies

Nonpositivity results due to our particular choices of the exposure to study and our confounder(s) of interest. In our hypothetical cohort, the exposure could not occur off work, and our confounder happened to be employment status. In that case, nonpositivity is assured. However, occupational epidemiology is not limited solely to the study of exposures that can occur at work, and employment status is not always a confounder of interest. There may be specific scenarios in which inverse probability weighting will succeed in occupational studies. Further, marginal structural models should be viewed as distinct from inverse probability weighting, and alternative strategies to fit a marginal structural model may be of use.

Control for bias due to movement out of dirty jobs or areas Consideration of healthy worker survivor bias has generally been confined to analyses in which illness may cause attrition from the workforce under study. However, the same types of bias can occur within a workforce of individuals that remain employed. For example, Christiani and colleagues examined the relationship between lung function and long-term exposure to endotoxin and cotton dust in a cohort of cotton textile workers (*Christiani et al.* (1999)). The authors observed feedback between lung function (measured by FEV_1) and dust exposures. That is, working in high exposure areas influenced changes in FEV_1, and workers with low FEV1 tended to move to lower exposed jobs. In principal, inverse probability weighting could have been used to estimate the association between cotton dust exposure and FEV_1 using these data, while controlling for confounding that occurs due to changes in work area. Such confounding could occur if, for example, workers with asthma tended to avoid higher exposure work areas.

This type of analysis may be best suited for settings in which workforce attrition over follow-up is low and the effects of exposure do not have long latent periods. Potentially, use of inverse probability of censoring weights (*Cain and Cole* (2009)) could

be used to appropriately deal with workers leaving employment, as speculated previously by *Joffe et al.* (2012). However, the use of such weights has not performed well in practice due to a strong underlying assumption that one has measured all covariates that could confound the association between leaving employment and the outcome of interest (*Picciotto et al.* (2013)).

Returning to our hypothetical example, imagine now that when workers left their jobs ($L_2 = 0$), they moved simply to another job at which they could be exposed. The data shown in Table 3.7 represent our original data, had we collected data on exposure among those who had left work. This results in a change in four cells of our table: a reallocation of individuals from the strata in which $L_2 = 0, X_2 = 0$ to the strata in which $L_2 = 0, X_2 = 1$. The mean of the stabilized weights is 1.0 in these data and our pseudo-population is the same size as the original population. Marginal structural models fit using inverse probability weighting yielded an unbiased mean difference of 18, while the crude and L_2-adjusted estimates are biased (mean difference = 17.8, 12.8, respectively, Table 3.8).

Control for bias using MSMs without IP weighting Marginal structural models fit using inverse probability weighting are attractive for their simplicity. However, the parametric g-formula (*Robins* (1986); *Keil et al.* (2014a)) allows one to fit a marginal structural model as well. The parametric g-formula is a tool used to estimate the distribution of the outcome under interventions on the exposure. This method can be used when there is nonpositivity. In inverse probability weighting, there is no extrapolation to empty strata strata with a weight (and Pseudo-N) of zero are not included in the marginal structural model. In contrast, the parametric g-formula can be used to extrapolate results to empty strata of exposure and covariates. In our example, the parametric g-formula can yield unbiased estimates of the dose response as long as the effect of increasing exposure by one unit is homogeneous between the employed and

Table 3.7: Mean outcome ($E[Y]$), stabilized inverse probability weights (SW^x) and pseudo-population cell sample size by strata of exposure (X_1, X_2) and employment status (L_2) in a hypothetical occupational cohort in which exposure is measured off work (and positivity holds)

X_1	L_2	X_2	$E[Y]$	N	SW^{x**}	$pseudo-N^*$
0	0	0	35	105	0.81	85
0	0	1	53	95	1.21	115
0	1	0	11	100	1.20	120
0	1	1	29	180	0.89	160
1	0	0	43	10	2.50	25
1	0	1	61	390	0.96	375
1	1	0	19	20	0.25	5
1	1	1	37	60	1.25	75
Overall**				960	1.00	960

* Pseudo-population cell sample size ($SW^x N$) rounded to nearest integer
** $SUM(N)$, $SUM(pseudo-N)$, $MEAN(SW^x)$

Table 3.8: Results of analysis on synthetic data from table 2 and table 3

Nonpositivity	Analysis	MD**	%Bias†
Yes	Crude	8.0	-56
	L adjusted	11.0	-38
	IPW-MSM	11.3	-37
	G-formula-MSM	18.0	0
No	Crude	17.8	-1
	L adjusted	12.8	-29
	IPW-MSM	18.0	0
	G-formula-MSM	18.0	0

* Data shown in 3.6 (nonpositivity) and 3.7 (positivity holds)
** Mean difference in the outcome per unit of exposure, the true value is 18
† %Bias =100*(estimate-truth)/truth

the unemployed.

The parametric g-formula uses predictive models and Monte Carlo sampling to generate a set of pseudo-data that represent the cohort under study, had we intervened on exposure. The outcome distribution in these pseudo-data is interpreted as the standardized outcome distribution under the intervention. A marginal structural model for our example fit with the g-formula is simply a model for the outcome that includes exposure as the only regressor, fit to pseudo-data in which the intervention is to randomize exposure.

Using the data shown in Table 3.6 in which nonpositivity occurs, the parametric g-formula algorithm given in appendix 2 yielded an unbiased mean difference of 18 (Table 3.8). For the data shown in Table 3.7 in which positivity holds, this method yielded an unbiased mean difference of 18, as well. The parametric g-formula treats the empty cells like sparse data, and smooths over them. Thus, it is not subject to the same bias from nonpositivity as inverse probability weighting.

Control for bias when exposure occurs off work A key assumption behind Robins and colleagues recommendation to avoid marginal structural models in occupational studies was that exposure could not occur off work. However, this assumption may rarely hold in practice. The agents of interest that motivate occupational cohort studies often are found outside the workplace. This can occur because these agents are transported from the worksite (e.g. on workers clothes or sold as a consumer product) or because the agents are not implicit characteristics of the worksite (e.g. particulate matter or radon exposure).

Nonpositivity arises in studies of some occupational exposures only as a function of our inability to capture exposure that occurs outside of employment (or because the investigator assumes that exposure does not occur outside of employment). In a population in which exposure could occur (and was measured) off work, the inverse

probability weights in the strata $X_2 = 1, L_2 = 0$ would no longer be undefined.

In some settings it is infeasible to continue follow-up for exposure measurements of employees who have left work. For some occupational exposures, however, population estimates of non-occupational exposure are available. These population level estimates may be useful to impute individual, off-work exposures, thus reducing or eliminating bias from nonpositivity. One might be willing to accept some misclassification of exposure (by using imputed exposures) in order to reduce bias due to confounding and nonpositivity.

To illustrate, consider an example from an occupation cohort of uranium miners, who are exposed to high levels of radon gas (*Schubauer-Berigan et al.* (2009)). Details about population characteristics, occupational radon exposure, and employment time have been reported previously (*Keil et al.* (2014b)). Occupational exposure to radon was estimated for workers during their period of employment. Residential radon exposures were not measured for the members of this cohort, but radon is a ubiquitous exposure that occurs in any enclosed space.

We imputed residential radon exposures for all person years under observation and added these radon exposures to occupational exposures to estimate total annual radon exposure. We evaluated results over a range of possible average radon exposures consistent with publically available estimates (*Price et al.* (2011)). We estimated the inverse probability weights using a multinomial logistic model for deciles of annual exposure (as in *Naimi et al.* (2014b)) adjusted for baseline covariates (race, birth cohort, year of hire, and mining related radon exposure and years employed as a uranium miner at baseline) and time varying covariates (prior annual exposures for 1,2,3,4,5, 6-9 years and cumulative exposure accrued 10 years prior and duration of employment since baseline). For a single set of imputed exposures, there is no guarantee that total radon exposure will be free from nonpositivity due to chance (i.e. we must observe individuals in each decile of exposure in all strata of covariates). To ameliorate this problem,

we created 200 copies of the original cohort, with unique imputations of residential exposures, and analyzed these copies in a pooled analysis.

Over a range of possible residential exposures, the hazard ratio varied from 1.0 at large exposures to about 1.08 at smaller imputed residential exposures (Figure 3.4, top panel). The mean stabilized inverse probability weights were close to 1.0 for exposures with a log-geometric mean less than -0.5, but the weights were very large for imputed residential exposures with log-geometric means greater than -0.5, likely due to non-positivity (Figure 3.4, bottom panel). The hazard ratio for the marginal structural model appeared to be biased downward relative to the crude hazard ratio of 1.05 for values of residential exposures that yielded weights with a mean close to 1.0.

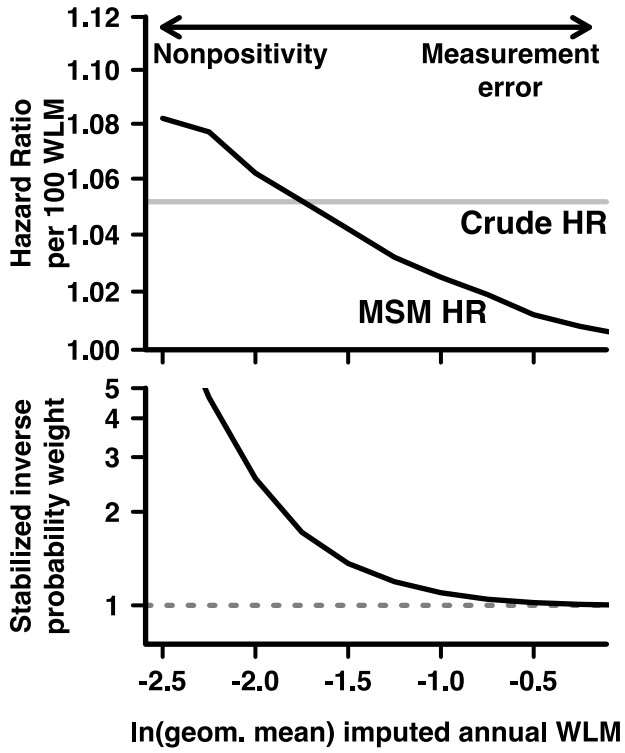

Figure 3.4: Bias tradeoff between nonpositivity and measurement error in the analysis of miner data; Top panel) Hazard ratio per 100 working level months from marginal structural Cox model (MSM) by imputed log-geometric mean of residential exposures, with grey reference line at the crude HR with no exposure imputation; Bottom panel) Mean inverse probability of exposure weights used in marginal structural Cox models by imputed log-geometric mean of residential exposures

3.2.8 Discussion

Occupational studies often involve settings of repeated or protracted exposure, which introduce potential problems due to time-varying confounding affected by prior exposure. Marginal structural models should not be eliminated from the toolbox of occupational epidemiologists, provided that they are used with care.

For a repeated binary exposure, Hernán et al define the effect measure from a marginal structural model as the ratio or difference in the outcome distribution at any time t had all subjects been continuously exposed compared with the outcome distribution at time t had all subjects remained unexposed (*Hernán et al.* (2000) p553). Effect measures from such a model may also be interpreted as the ratio or difference in the mean outcome per unit increase in cumulative exposure, which is the same interpretation given to more familiar regression models. Thus, marginal structural models allow easy contextualization of results within the larger literature of regression-based dose-response estimates. Inverse probability weighting is also a relatively simple estimation procedure.

Likely because of these attractive features, marginal structural models fit with inverse probability weights have seen some use in the occupational literature. However, modifications were necessary to prevent nonpositivity. Dumas et al utilized marginal structural models to estimate the effect of occupational exposure to asthmagens on asthma expression in a clinical population (*Dumas et al.* (2013)). This approach is similar to our approach, in that they do not limit exposure estimates to a single workforce. However, Dumas et al do not consider off-work exposures to asthmagens and do not account for active employment status, a likely source of confounding in their population (*Le Moual et al.* (2008)), which is a mix of employed and unemployed individuals. Similarly, Thygesen et al use marginal structural models to estimate the association between employment as an electrician in years 1 and 3 on mortality at the end of year 3 (*Thygesen et al.* (2011)). The confounder of interest was employment in year 2. The

authors avoid issues of nonpositivity by ignoring possible confounding by exposure in year 2 and employment in years 1 and 3. Both sets of authors reported mean weights close to 1.0, suggesting they had little or no bias from nonpositivity. There would almost certainly be bias from nonpositivity had both studies included the noted confounders, however.

These examples and our example with the miner data suggest that there is a tradeoff between bias from nonpositivity and bias from any approach to avoid it. We can identify the direction of the bias in our miner example by comparing it to a previous analysis in which we controlled healthy worker survivor bias using a method that is not subject to nonpositivity (*Keil et al.* (2014b)). In that case, the bias from confounding by employment history was downward, indicating that our current example does not result in a net reduction in bias once we reduce or eliminate nonpositivity. As in the analyses by Dumas et al and Thygesen et al, we will not always have this prior information, however. To choose the solution with lower overall bias, we would first need to quantify the magnitude of bias from nonpositivity.

Petersen et al developed a method to detect and quantify bias from nonpositivity (*Petersen et al.* (2012)). The authors approach was to compare a marginal structural model fit using inverse probability weighting to a marginal structural model fit using the parametric g-formula. If both models are correctly specified, the difference between them is the bias from nonpositivity. This method is of limited practical use when nonpositivity is guaranteed. Once one has gone through the effort of using the parametric g-formula, there is hardly any need to use inverse probability weighting. Instead, for occupational studies the parametric g-formula can be used in general, and inverse probability weighting can be used when one can reasonably assume that nonpositivity will not occur.

Using the parametric g-formula allows us to retain the simple interpretation of the marginal structural model. Using the same miner data used in our analysis, Edwards

et al recently used the g-formula to estimate the effect of occupational radon exposure on lung cancer mortality (*Edwards et al.* (2014)). In contrast to our approach, the authors did not estimate a simple dose-response, however. The authors proposed estimating the change in the incidence of lung cancer under different caps on occupational exposure levels. To estimate the effects of these interventions, the authors had to develop models for exposure, employment status, and lung cancer mortality. This suggests several potential difficulties with applying the parametric g-formula to cohort studies. The need for multiple models increases the potential bias due to model misspecification and contrasts with the relative simplicity of inverse probability weighting. Nonetheless, fitting marginal structural models with the parametric g-formula remains widely applicable in occupational studies, in spite of the paucity of applications in the occupational literature (*Cole et al.* (2013); *Edwards et al.* (2014)).

An alternative to marginal structural models, structural nested models, have been used to estimate dose-response parameters in occupational studies when nonpositivity is a concern (*Naimi et al.* (2014a), §3.1 of the current manuscript). These models, while appealing with respect to control of confounding by time-varying covariates, are more technically challenging and can suffer from convergence problems (*Joffe* (2001)). Structural nested models are limited with respect to the effect measures that can be estimated, necessitating further assumptions to translate effect measures to more familiar ones (*Chevrier et al.* (2012)). Structural nested models may be more useful than marginal structural models, however, when there is nonpositivity and one is concerned about model misspecification in the parametric g-formula.

3.2.9 Conclusions

Bias of exposure-outcome relationships can come from many sources, and improvements in study design and implementation can minimize bias from many of these sources. In occupational studies, however, we are limited to the extent we can prevent

bias due to time-varying confounders. This limitation necessitates the use of analytic methods to control this source of bias. As our synthetic example shows, conventional, regression based approaches can fail to control confounding, and in some cases can increase bias. Marginal structural models should not be ruled out as an alternative to such methods in occupational studies. Under a limited set of occupational settings, inverse probability weighting can be used to estimate dose-response parameters in a marginal structural model. This supplements the possible use of marginal structural models noted by Joffe to estimate other parameters of interest in occupational studies (*Joffe* (2012)). Our cautionary example, and limited use in the literature shows that certain strategies to avoid nonpositivity may result in an increase, rather than a decrease in bias. More generally, marginal structural models can be used in occupational settings when they are fit using the parametric g-formula. Marginal structural models are rarely used in occupational studies, but their proper use should complement that of structural nested models when addressing time-varying confounding.

3.2.10 Appendix

Why inverse probability weighting fails when there is non-positivity To understand why inverse probability weighting fails when there is nonpositivity, note that the technique was originally developed to make inference from survey samples (*Horvitz and Thompson* (1952)). Weighting relies on treating the observed population as though it were sampled from a target population, with sampling fractions that are proportional to $(SW^x)^{-1}$. In our case, the target population is a hypothetical population in which covariates are balanced over exposure. In some strata of the covariates of the observed population, there are zero individuals, so there are no individuals to weight-up to represent individuals from these strata in the target population. Thus, the pseudo-population is not a good substitute for the target population, and bias results. This is analogous to the problems that would arise with making population inferences from

US census data if, say, no individuals under the age of 30 were surveyed.

Note that, the weights in our hypothetical population from Table 2 are not incorrect, even though they do not correspond to weights referring to a target population. Further, the software we use may not help detect nonpositivity. For example, we estimated the weights for Table 3.6 using the LOGISTIC procedure in SAS 9.3. This approach yielded a warning message in the log regarding quasicomplete separation of the data points, which is an indication of empty cells in a tabular analysis (*Albert and Anderson* (1984)). In contrast, the same task done using the GLM function of R 3.1.0 yielded no warning. Weight estimates from both programs were identical. If we treat this as a problem of sparse data by, for example, adding 0.5 to each cell for weight estimation, the bias will persist even if the warning messages stop (see Appendix B.1 for an example of this).

CHAPTER IV: CONCLUSION

4.1 Overview

Healthy worker survivor bias has had a long history in the occupational literature as a potential problem for estimating the impacts of occupational exposures. This bias was detected in frequently surprising settings in which cumulative exposure was associated with improving health, contrary to what was known from experimental evidence or studies done in other settings. A parallel, and older observation of healthy hire bias, was that working populations selected for study often had lower disease and mortality rates - this observation occurred independent of exposures because the only aspect under study was employment. While there have been substantial efforts to distinguish these two phenomena from each other, they can also been seen as two distinct outcomes from a rather simple principle: on average, employed people are healthier than unemployed people.

The detection of healthy worker survivor bias often relied on studying exposures with small enough effects that the bias would produce non-sensical results. Some work has been done to show that unexpected curvature in the dose-response curve can result from this bias. In the radon literature, however, no occupational studies exist in which radon appeared to be beneficial, and there are several alternative explanations for curvature in what should be a linear dose-response. Thus, there are no useful criteria under which we would expect healthy worker survivor bias of the radon-lung cancer dose-response. A unifying principle "employed people are healthier than unemployed people" suggests evidence of this bias in such studies, even though none of the classical criteria have been met. There are many examples of apparent healthy hire bias

in the miner literature, which may be evidence of this principle. The primary aims of this dissertation were to quantify healthy worker survivor bias in the Colorado Plateau and to suggest possible new approaches to quantifying and ameliorating such bias. To address this problem, we sought to apply two models from the causal inference literature: structural nested accelerated failure time (SNAFT) models and inverse probability weighted marginal structural models. We further sought to elaborate on previous authors' assessments about when inverse probability weighting will and and will not be a desirable approach in occupational studies.

4.2 Summary of results and future directions

In the current manuscript, we have shown that structural nested models are useful for estimating cumulative dose-response metrics in occupational data subject to healthy worker survivor bias. Further, we suggest that healthy worker survivor bias may be present in at least one miner study that has been used to estimate population risk from radon exposure. We show that inverse probability weighted marginal structural models may not be useful in miner studies to estimate this dose-response, but we speculate that these models may be useful in other areas of occupational epidemiology. The results of our analysis suggest that the population impact of residential radon exposures may be underestimated when they are based on relative rate parameters from miner studies, and we propose that reanalysis may warranted.

To close, we speculate on some possible directions in which the use of g-methods may be useful for occupational epidemiologists, and we suggest some improvements that may make estimation of such models more robust.

4.2.1 Moving past cumulative exposure metrics in g-methods

Cumulative exposure is a summary metric that accounts simultaneously for exposure duration and average exposure intensity. It is a useful metric that has proven to provide robust predictions of exposure rates in several settings (*Thomas* (2014)) and results in effect measures that are simple to communicate. The work of *Langholz et al.* (1999),*Richardson and Ashmore* (2005), *Richardson* (2009b), and *Richardson et al.* (2012) has suggested another route of modeling using an aggregated exposure that is a weighted sum of time specific exposures, where the weights can account for factors such as latency or exposure intensity. This more general approach may be possible using structural nested models by considering models of the form

$$T^{\tilde{\mathbf{0}}} = \int_{k=0}^{T} \left(1 + \int_{u=0}^{k} \psi X_u w_u \right) d\,k$$

Where w_u is some function that weights the aggregate exposure metric $\bar{X}_k = \int_{u=0}^{k} X_u$ by possibly time varying factors. As in *Richardson et al.* (2012), if we let w_u be of the relative effect of age at exposure $w_u = \exp(\gamma \times \text{age}_u)$, such an approach can be used to estimate the modification by age at exposure that does not force the use of the cumulative exposure metric and thus may cohere better to our understanding of disease processes from acute exposures (e.g. the study of the atomic bomb survivors *Preston et al.* (2007)). This approach is well suited for structural nested models because of the selective ignorability assumption (*Joffe et al.* (2010)). This implies that we could fit such a model only to the data in which exposure occurs (as in the current analysis), thus reducing the need for an "age at exposure" metric that is summarized across many time periods. The difficulties encountered in the main analyses with fitting models for modification of the time ratio suggest that this approach might be best attempted using pooled data from multiple cohorts.

4.2.2 Unifying g-methods with classical risk projection models

One strength of the g-methods (g-estimation of structural nested models, the parametric g-formula, inverse probability weighted marginal structural models) is that they can be easily used to compare outcome distributions under different exposure regimes. For example *Chevrier et al.* (2012) used SNAFT models to estimate the cancer mortality under an exposure regime "expose to metalworking fluids for 5 years." Similarly *Cole et al.* (2013) and *Edwards et al.* (2014) used the parametric g-formula to estimate the effects on mortality of different occupational limits on asbestos or radon exposure. These approaches are ideal in that they represent estimates of public health interventions and are ideally suited for informing decision makers.

Another approach is to transport the relative rate estimates from miner studies to estimate the population risk from residential radon. The typical approach from the BEIR VI committee and the USEPA (among others) is to use miner data to inform estimates of the impact of residential exposure to radon (*NRC* (1999); *USEPA* (2003)). Their approach is to take estimates of the relative rate parameters from the miner studies and, after accounting for differences in the expected equilibrium of radon with its progeny and differences in respiration, multiply these relative rate parameters by the baseline cancer rates under various intensities of residential exposure. Using life tables, these expected rates are compounded over the life-course of a standard population to estimate lifetime risks for the general population. G-methods are suited for such an analysis that may generalize this approach to relax some of the assumptions in such an analysis.

Using the g-formula given in Equation (1.8) as an example, if we have a standard population in which we have measures of $\boldsymbol{V}$, $\bar{\boldsymbol{L}}_{\boldsymbol{k}}$, and $\bar{\boldsymbol{X}}_{\boldsymbol{k}}$ and an occupational cohort in which we have estimated $Pr(L_k|\cdot), Pr(V|\cdot), Pr(X_k|\cdot)$, and $Pr(D_k|\cdot)$, it is a simple extension of available formulations of the g-formula (e.g. *Keil et al.* (2014a)) to estimate the incidence of disease in the standard population by using the probability models

generated in the occupational cohort. This is a form of indirect standardization to estimate the impact of exposures in the general population, informed by occupational data.

4.2.3 Contrasting structural approaches for exposures whose effects may be persistent

The approach taken in the current analysis to summarize the effect of radon exposure on mortality is to consider an approach in which exposure at one time can influence the outcome at multiple time-points. This approach is typical of regression models in which some aggregated exposure is included a model for the hazard or mortality rate.

The way in which data are analyzed is to split up the duration of follow-up into discrete periods of time. For example, a version of the relative rate model similar to models shown in §1.3 can be expressed as

$$\lambda(k|\bar{X}_k;\beta) = \lambda_{0k}\, e\, x\, p(\beta\, \bar{X}_k)$$

can be expressed in terms of its partial likelihood

$$\mathscr{L}_p(\beta) = \prod_{k=T_i} \frac{\exp(\beta\, \bar{X}_{ik})}{\sum_{j\in\mathscr{R}_i} \exp(\beta\, \bar{X}_{jk})}$$

Which is evaluated at the event times T_i. It is necessary to include cumulative exposure (or some time integrated exposure $\bar{X}_k$) if we believe that exposure has persistent effects. To see this, note that if we substitute $\bar{X}_k$ with current exposure X_k, the exposure at some previous time is not accounted for in the partial likelihood. Such a model may be useful if we believe that some agent has very prompt, but temporary effects on an outcome of interest. The issues remain identical if we fit a marginal structural (MS) Cox proportional hazards model, in which case the partial likelihood is augmented by

separate weight terms for the cases and the risk set members (cf *Cole and Hudgens* (2012)).

Contrast this model with the SNAFT model used in the current analysis as well as by *Hernán et al.* (2005) and *Naimi et al.* (2014a):

$$T^{\bar{\mathbf{0}}} = \int_{k=0}^{T} \exp(\psi \bar{X}_k) d\,k.$$

Again, we include cumulative exposure in the model. However, note that if we replace $\bar{X}_k$ with current exposure X_k, the exposure will still have an effect on the time ratio even if it occurs long before the outcome. This happens because, unlike the marginal structural Cox model, the SNAFT model includes exposure in the integration term. Thus, the SNAFT model can include only current exposures and still detect persistent effects of exposure, whereas the MSM cannot.

As an example, consider an exposure that affects a disease with a 5 year latency period and has persistent effects - assume this period is the same for every member of the population. If current exposure is used in the model, a MS Cox model using only current exposure will fail to detect any effect of exposure, while the SNAFT model will detect this effect in a large sample.

The difference between these two approaches has implications for how exposures are modeled in SNAFT models. A different set of authors (*Chevrier et al.* (2012); *Neophytou et al.* (2014); *Picciotto et al.* (2014)) has previously used SNAFT models for occupational agents with potentially persistent effects (metalworking fluids). These authors fit SNAFT models that included only current exposure in the model, in contrast with our approach, but they still observed positive associations with outcomes that occurred long after the exposure ceased (note that this does not exclude bias as a source of the observed association). These authors' approach may seem unfamiliar to epidemiologists more familiar with models for the rate or hazard, but highlights a key dif-

ference between such models and accelerated failure time models. Accelerated failure time models allow for current exposure at any time in the life course to influence the final outcome, whereas exposure must be aggregated across time in a rate or hazard model to allow for persistent effects.

Our approach is desirable in that we are able to easily compare parameters from our SNAFT model with more standard approaches, whereas the approach of *Chevrier et al.* (2012) required comparing potential outcomes $T^{\bar{0}}$ with potential outcomes under 5 years of exposure using a Cox model. This extra step was necessary, because their time ratio compared the potential outcomes under "unexposed" with "always exposed" whereas ours compares "unexposed" with "kept at a constant cumulative exposure of 100 working level months." Our intervention may be more realistic intervention for an occupational setting. However, cumulative exposure may not be the metric that yields the best prediction of public health benefits.

Future work should contrast these two approaches using SNAFT models. There are likely implications or for estimating time-related aspects of exposure effects, such as when considering effect measure modification by age at exposure. One possible avenue to approach this work is the instantaneous-rate SNAFT model discussed in sections 7 and 8 of *Robins* (1998). These models allow for effects on the failure time (time of death, time to disease) of exposures that occur only within a certain window of time before the event. If the exposure does not occur within that time, it is assumed to have no effect on the outcome. The instantaneous-rate SNAFT model may coincide better with biologic knowledge in some areas of cancer research, in which cellular repair processes may, over time, reduce the cellular damage from environmental insults such as radon.

4.2.4 Flexible approaches to exposure modeling with structural nested models

In sensitivity analysis for the SNAFT models shown in Appendix A, we attempted to fit several alternative models for the exposure. The estimating equations approach we took utilizes information in the "novel" part of the exposure. This is the variation in an individual's exposure from the predicted exposure. The approach contrasts this with the potential outcome (or, in more efficient versions of the estimating equations, the "novel" part of the potential outcome, as well). Thus, there is a need to find an exposure model that both does not poorly fit the expected exposure, but also does not over-fit the exposure so as to remove some of the "novelty" from the exposure. We have taken a parametric approach to modeling exposure, which is especially prone to poor predictions. Thus, it may be of use to find flexible prediction models that can avoid over-fitting.

We attempted to address this potential shortcoming in a sensitivity analysis in which we fit an exposure model using generalized additive models or discrete approximations with a multinomial model. However, these approaches did not lead to convergence of the SNAFT model. One possible future avenue is to use Super Learner, which is a cross validated prediction method that utilizes model averaging over a library of predictive models (*van der Laan et al.* (2007); *Polley and van der Laan* (2010)). Super Learner could be used to predict the exposure in a way that avoids the restrictive assumptions of parametric models but is nonetheless useful to avoid over fitting. Because the estimating equations approach requires only a single prediction model for the algorithm, this approach would not lead to a substantially longer time for analysis relative to our approach.

APPENDIX A: FURTHER DETAILS OF STRUCTURAL NESTED MODEL ANALYSIS OF THE COLORADO PLATEAU URANIUM MINER COHORT

A.1 Covariates and notation used in the analysis of the Colorado Plateau uranium miner data in Appendix A

Table A.1: Notation used to refer to quantities in the Colorado Plateau Uranium Miner (CPUM) data

Symbol*	Interpretation	CPUM Variables
X_k	Primary exposure	Monthly radon exposure in 100 working level months, lagged 5 years
$\bar{L}_k$	Time varying confounders	Employment status (yes or no), unlagged
$\bar{\boldsymbol{X}}_{\boldsymbol{k}}$	History of exposure	Separate terms for annual exposure lagged 1, 2, 3, 4, 5, and 6-10 years, and cumulative exposure 10 lagged 10 years
$\boldsymbol{V}$	Baseline covariates	Race (white, other), smoking status at 1985 interview (current, former, never), birth cohort (<1890, 1890-1899, 1900-1909, 1910-1919, 1920-1924, 1925-1930, 1930-1934, 1935-1939, >1939), prior radon exposure from hard rock mining (working level months), baseline radon from uranium mining, baseline number of years employed in uranium mining, year of hire (cubic polynomial)
D_k	Outcome of interest in month k (yes or no)	Date of death from lung cancer, other causes
T	Time to outcome of interest	Date of death from lung cancer, other causes
T^0	Potential time to outcome of interest if never exposed	Not identified in data

* Vectors are shown in **bold** and scalars shown in regular typeface. Time varying quantities are subscripted by month k.

A.2 Technical details: g-estimation of a structural nested accelerated failure time model

A.2.1 G-estimation algorithm accounting for censoring by end of follow-up and competing risks

We now describe further the structural nested model used to estimate the time ratio per 100 WLM of exposure in the main text. We note that our choice of the model for the effect of exposure (repeated from the text) is shown in A.1.

$$T^{\bar{0}} = m + \int_{m}^{T} (1 + \psi \bar{X}_k) d\,k \tag{A.1}$$

This SNAFT model differs from typical published examples (such as that shown in modelA.2 and e.g. *Witteman et al.* (1998); *Chevrier et al.* (2012); *Naimi et al.* (2014a)).

$$T^{\bar{0}} = m + \int_{m}^{T} exp(\psi \bar{X}_k) d\,k \tag{A.2}$$

However, we show that our model is a valid form for a SNAFT model. Equation A.1 can be expressed equivalently using the function that relates the observed failure times to the potential failure times as:

$$T^{\bar{0}} = m + \int_{m}^{T} (1 + \psi \bar{X}_k) d\,k = h(T, \bar{X}_k, \psi) \tag{A.3}$$

Thus, the function $h(T, \bar{X}_k, \psi)$ takes as its input the observed failure time, the exposure of interest, and the value of 1-TR and yields the potential failure times we would observe under no exposure.

Robins notes that the function $h(T, \bar{X}_k, \psi)$ should satisfy three criteria:

1. identity: $h(t, \bar{X}_k, \psi) = t$ if $\bar{X}_k$ is always 0 from m to t

2. consistent with null-hypothesis: $h(t, \bar{X}_k, \psi = 0) = t$ so that $\psi = 0$ represents the null hypothesis of no effect of exposure on the time of death

3. monotonicity: $h(t, \bar{X}_k, \psi) > h(u, \bar{X}_k, \psi)$ if $t > u$

The model shown in (A.3) meets conditions 1 and 2 by noting that if either ψ=0 or $\bar{X}_k$ is always 0, that the function reduces to $m + \int_m^t 1 dk = t$. In the case of radon and lung cancer, (A.3) meets the third criteria by the linear-no threshold assumption underlying the carcinogenic effects of radon. Practically, this implies that the TR will always be positive. In the case of parameter estimates that imply a protective effect of radon, we need only assume that $\psi * MAX(\bar{X}_k) > -1$, which guarantees monotonicity. In the miner data, we observed a maximum cumulative exposure of around 60*100 WLM, implying that we would observe practical violations of monotonicity for estimates of the TR < 0.983. Our analyses uniformly indicate TR> 1, so this is not a concern in the miner data.

Censoring occurs in our analysis of the CPUM under three separate mechanisms: individuals are considered censored if they a) are lost to follow-up prior to death, b) die from a cause other than lung cancer or c) survive until the end of the study on 31 December 2005. Following other analyses of the CPUM, we consider a. and b. to be equivalent and we assume that there are no unmeasured risk factors for the censoring event. Under censoring events a, b or c, where C is the time at censoring, analysis of time to event data typically considers estimation using $Z \equiv MIN(T, C)$ and $\Delta \equiv I(T < C)$, where $I(\cdot)$ is the indicator function taking on values 1 or 0.

For the SNAFT model, we consider similar variables $Z^{\tilde{\mathbf{0}}} \equiv MIN(T^{\tilde{\mathbf{0}}}, C^{\tilde{\mathbf{0}}})$ and $\Delta^{\tilde{\mathbf{0}}} \equiv I(T^{\tilde{\mathbf{0}}} < C^{\tilde{\mathbf{0}}})$. Redefining the administrative censoring time C as $C^{\tilde{\mathbf{0}}}$ is necessary because of one of the following reasons 1) if exposure is harmful ($\psi > 0$), then we would expect that some of the events observed under no exposure would be unobserved if, in fact, the individual had been unexposed (i.e. life would be extended past C by remaining unexposed) and $C^{\tilde{\mathbf{0}}}$ should allow potential failure times to be observed only if they

would have been observed under the minimum exposure; or 2) if exposure is beneficial ($\psi < 0$), then those with the shortest potential times under no exposure will (in general) be the highest exposed, so $C^{\bar{\mathbf{0}}}$ should allow potential failure times to be observed only if they would have been observed under the maximum exposure. If exposure has no effect on survival ($\psi = 0$), then censoring times do not need to be adjusted.

Since radon is an established risk factor for lung cancer, we only need consider this calculation when $\tilde{\psi} > 0$. If we define the age at start of follow-up as m, then a possible function for $C^{\bar{\mathbf{0}}}$ under our SNAFT model is

$$C^{\bar{\mathbf{0}}} = m + (1 + \tilde{\psi}\ MIN(\bar{X}_k))(C - m) \tag{A.4}$$

In other words, since the minimum possible cumulative radon exposure is 0, $C^{\bar{\mathbf{0}}} = C$. This function results in wider confidence intervals than the optimal functions for $C^{\bar{\mathbf{0}}}$ discussed by Robins and Tsiatis (1992), but it is much simpler to calculate.

Because exposure may affect how many events are observed under administrative censoring, $C^{\bar{\mathbf{0}}}$, rather than C is used to define the set $(Z^{\bar{\mathbf{0}}}, \Delta^{\bar{\mathbf{0}}})$ so that the identifying assumption of no unmeasured confounding (see §C.3)an be written as

$$(Z^{\bar{\mathbf{0}}}, \Delta^{\bar{\mathbf{0}}}) \perp X_k | \bar{L}_k, \bar{X}_{k-1}, V_0, T > k \tag{A.5}$$

Under the assumption of no unmeasured confounding, we can test (A.5) using the methods given in A.2.3, which forms the basis of g-estimation.

To adjust for selection bias due to possibly informative censoring due to deaths from other causes or loss to follow-up. We utilized inverse probability of censoring weights, where, at each time k, we estimate the weight W_k as

$$W_k^c = \frac{Pr(C_k = 0 | \bar{X}_{k-1}, V_0, C_{k-1} = D_{k-1} = 0)}{Pr(C_k = 0 | \bar{L}_{k-1}, \bar{X}_{k-1}, V_0, C_{k-1} = D_{k-1} = 0)} \tag{A.6}$$

The weight model is fit to data before 12/31/2005, since we account for censoring

due to end of follow-up as described above. While we do not include the effects of smoking in our exposure model, we include smoking status as of 12/31/1985 (never, former, <1 pk/day, 1 pk/day, >1pk/day) in V_0 to allow for the effect of smoking on deaths due to other causes.

A.2.2 Description of exposure model used for g-estimation algorithm

Adjusted model The estimating equation relies on knowing (or estimating) $p(X_k)$, the time specific expected value of the exposure, conditional on the covariates included in (3). Because radon is recorded as a quantitative variable (and because we are interested in the dose-response of the radon-lung cancer relationship), the expected value of radon exposure at age k is sensitive to the choice of parametric model for the mean. Preliminary analysis indicated that monthly exposures while working as a miner approximately follow a zero-inflated, right skewed distribution. We define the predicted exposure as

$$p(X_k) = p\,r(X_k > 0|\cdot)E(X_k|X_k > 0,\cdot) \tag{A.7}$$

Where $\cdot$ is the set $(\bar{L}_k, \bar{X}_{k-1}, V_0, T > k)$, $p\,r(X_k > 0|\cdot)$ is estimated in a data-set of person-month records using pooled logistic regression over all person-months and $E(X_k|X_k > 0,\cdot)$ is estimated using a log-linear model for exposure fit to all exposed person-month records. Thus, the expected radon exposure at time k is the proportion exposed to any radon times the mean radon exposure, in strata of the predictors of exposure. We refer to this model as a "zero-inflated log-linear model," which we fit with a joint-likelihood similar to that shown in *Li et al.* (2011), but with a log-linear (e.g. the mean of the distribution is an exponential function of predictors, but the error term is considered to have a Gaussian distribution) component in place of the log-normal component. Using this general methodology, we fit other such "zero inflated" models, discussed

more in §A.4.1.

Unadjusted model SNAFT models do not yield a parameter that estimates the magnitude of healthy worker survivor bias. It is of interest to know the magnitude of this bias to understand the potential impact of this bias on the accuracy of current risk projection models. Therefore, we fit a separate "unadjusted" model that could be compared with the adjusted model. Namely, such a model can be derived under the assumption that there is no time-varying confounding of the radon-mortality association, which, more formally, can be stated as the assumption of no unmeasured confounding for cumulative exposure:

$$(Z^{\bar{\mathbf{0}}}, \Delta^{\bar{\mathbf{0}}}) \perp \bar{X}_k | V_0, T > k \tag{A.8}$$

Which suggests a model for the exposure given as:

$$p(\bar{X}_k) = p\,r(\bar{X}_k > 0 | V_0, T > k) E(\bar{X}_k | \bar{X}_k > 0, V_0, T > k) \tag{A.9}$$

Essentially, A.8 states the underlying assumption necessary to give a causal interpretation to a standard accelerated failure time model. However, by using a SNAFT model to estimate the acceleration parameter, we ensure that the models do not differ with respect to assumptions about the shape of the baseline survival function. As noted in 3.1, however, our approach requires fitting multiple exposure models. We opted for our approach cecause we are more confident in our ability to characterize the exposure distribution than in our ability to accurately parameterize the shape of the underlying survival function. Here, the underlying survival function is the distribution of survival times we would observe, had the cohort never been exposed to radon. This is not a quantity typically reported in analyses of miner data.

A.2.3 G-estimation using estimating equation methodology

In the SNAFT analysis from chapter 3.1, we estimated ψ using the g-estimation algorithm described by *Robins and Tsiatis* (1992) and in detail by *Hernán et al.* (2005) and in the appendix of *Joffe et al.* (2012). This algorithm uses a modified score function $U(\tilde{\psi})$ similar to that described in Chapter 6 in *Kalbfleisch and Prentice* (2002). The estimate $\hat{\psi}$ is the value of $\tilde{\psi}$ for which the score function is zero. This is analogous to maximum likelihood methods, in which the score function (the partial derivative of the likelihood function) will equal zero at a local maximum in the likelihood function. This approach is advantageous in that $U(\tilde{\psi})$ is easy to calculate and a full regression model is not needed for each candidate $\tilde{\psi}$.

To find $\hat{\psi}$, we used the estimating equation

$$U(\tilde{\psi}) = \sum_{k=m}^{min(T_i,k_{L_i})} \sum_{i=1}^{N} [(X_{ik} - p(X_{ik})]\Delta_{ik}^{0}(\tilde{\psi})W_{ik}^{c} \tag{A.10}$$

where X_{ik} is the exposure in month k for individual i, $p(X_{ik})$ is the expected monthly radon exposure (described in §A.2.2) in month k, given the exposure and covariate history and that the miner is still at risk for lung cancer. $\Delta_{ik}^{0}(\tilde{\psi})$ is the indicator of whether or not the individual is not censored by the end of follow up or a competing risk (0=censored, 1=experiences event of interest), under the candidate value for the time ratio -1 $\tilde{\psi}$. The weight W_{ik}^{c} is the inverse probability of censoring (due to competing risks) weight shown in Equation A.6. The solution $\tilde{\psi} = \psi$ is found by finding value of $\tilde{\psi}$ at which $U(\tilde{\psi}) = 0$ either through a grid search, a non-gradient-based optimization algorithm, or any optimization algorithm that calculates an empirical gradient and can be run with multiple initial values. Gradient-based optimizers can be used effectively when $U(\tilde{\psi}) = 0$ is a smooth function, such as if a cohort is followed to extinction, but due to our choice of estimating equation, our estimating equation is not a smooth function, as can be seen in Figures A.1 and A.2. Joffe et al. (2012) describe the

performance of several optimization routines. The lack of a smooth estimating function also implies that $U(\tilde{\psi}) = 0$ will not equal 0 for any value of $\tilde{\psi}$, instead the value closest to 0 is used. The exposure model for $p(X_k)$ can typically be fit using standard software for generalized linear models, but solving the estimating equation requires a separate procedure.

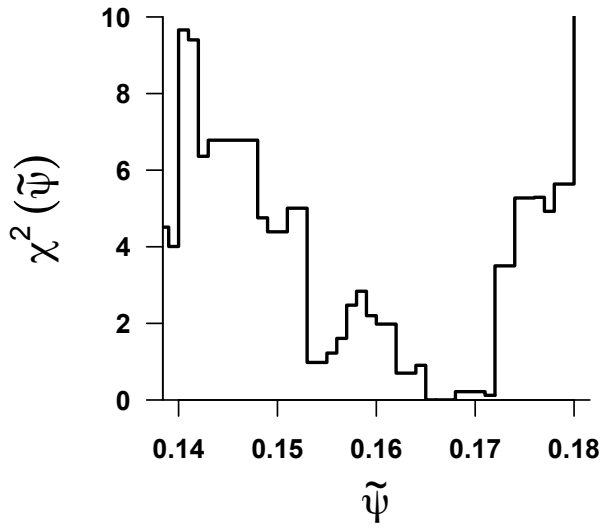

Figure A.1: G-function for SNAFT model for lung cancer mortality reported in §3.1

To estimate confidence intervals for $\hat{\psi}$, we again followed the approach outlined by *Joffe et al.* (2012). Briefly, this approach involves minimizing a χ^2 test statistic that can be calculated by

$$\chi^2(\tilde{\psi}) = U(\tilde{\psi})V(U(\tilde{\psi}))U(\tilde{\psi}) \tag{A.11}$$

Where $V(U(\tilde{\psi}))$ is the empirical variance of $U(\tilde{\psi})$, calculated as

$$V(U(\tilde{\psi})) = \sum_{k=m}^{min(T_i,k_{L_i})} \sum_{i=1}^{N} \left[[(X_{ik} - p(X_{ik})]\Delta_{ik}^0(\tilde{\psi})W_{ik}^c\right]\left[[(X_{ik} - p(X_{ik})]\Delta_{ik}^0(\tilde{\psi})W_{ik}^c\right]' \tag{A.12}$$

In large samples $\chi^2(\tilde{\psi})$ has an expected χ^2 distribution with degrees of freedom

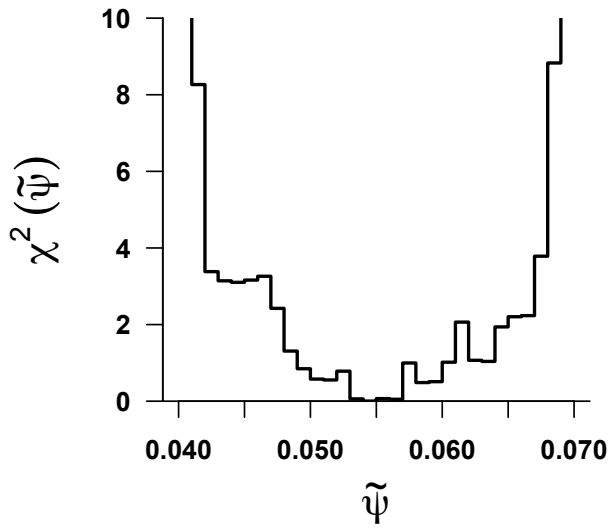

Figure A.2: G-function for SNAFT model for all cause mortality reported in §3.1

equal to the dimension of ψ. Since $U(\tilde{\psi})$ is a discrete function of potential survival indicators $\Delta^{\tilde{\mathbf{0}}}$, the minimum $\chi^2(\tilde{\psi})$ will not reach 0, $\chi^2(\tilde{\psi})$ will be a step function and may not be monotonic. Instead $U(\tilde{\psi})$ (and $\chi^2(\tilde{\psi})$) were calculated over a grid of potential $\tilde{\psi}$ values and the (mean) value(s) of $\tilde{\psi}$ for which $\chi^2(\tilde{\psi})$ was smallest were accepted as the estimate $\hat{\psi}$, the log time-ratio for a 100 working level month increase in cumulative radon exposure. We perform g-estimation using a grid search for models with a single ψ parameter and utilize an active-set algorithm with multiple start points for models with multiple ψ parameters. The full g-functions for the grid searches for our primary analyses with lung cancer and all cause mortality are shown in Figures A.1 and A.2.

For our primary analysis, we calculated 95% confidence intervals as the set of $\tilde{\psi}$ values surrounding the point estimate for which $\chi^2(\tilde{\psi}) < 3.84$. Our method yields conservative values for the 95% confidence intervals, but is considerably less computationally intensive than alternative methods, such as the bootstrap.

As noted above, a g-estimation algorithm can also be fit by adding $\Delta^{\tilde{\mathbf{0}}}$ to the exposure model, but this method is much more computationally intensive, since a model for exposure is fit for every value of $\tilde{\psi}$ in a grid search, whereas $U(\tilde{\psi})$ is generated us-

ing relatively fast matrix calculations (*Hernán et al.* (2005); *Joffe et al.* (2012)), and $U(\tilde{\psi})$ readily incorporates multi-dimensional $\tilde{\psi}$ (i.e. if we were interested in the effects of multiple exposures). Further, optimization routines may be implemented as an alternative to grid searches as a way to estimate ψ, though this approach may be problematic in practice (*Joffe et al.* (2012)).

A.3 Simulation studies: Fitting structural nested accelerated failure time models using g-estimation

To assess the success of our g-estimation algorithm in estimating causal parameters, We performed simulations under a data generating structure that produced observed data similar to that in the CPUM.

A.3.1 Data generation algorithm

Data were simulated according to a structural nested accelerated failure time model based on a modified version the algorithm described originally by Robins (1992) *Robins* (1992) and refined by Young and colleagues (2010) *Young et al.* (2010), which proceeds according to the following steps.

1. Generate a set of potential failure times $T^{\bar{0}}$ under a Weibull distribution with shape ρ and scale λ, $W(\rho,\lambda)$
2. Generate frailty variable $F = sqrt(T^{\bar{0}})$
3. Generate baseline versions time varying covariates (L_0) and exposure (X_0) which are, by convention 0
4. Start at $t = 1$, the end of the first time period (e.g. year) of the study
5. Generate static covariates, V, time varying covariates (L_k) and exposure (X_k)
   (a) V follows a Bernoulli distribution $B(p)$ where $p = Pr(V = 1) = expit(r_0 + f_f F)$
   (b) L_k follows a Bernoulli distribution $B(p)$ where $p = Pr(L_t = 1) = expit(g_0 + g_t\sqrt{t} + g_v V + g_{Cl}\bar{L}_{k-1} + g_l L_{k-1} + g_x \bar{X}_{k-1} + g_f F)$,
   (c) X_k follows a log-normal distribution $LN(\mu,\sigma)$ where $\mu = ln(E(X_t)) = a_0 + a_t\sqrt{t} + a_v V + a_{Cl}\bar{L}_{k-1} + a_l L_{k-1} + a_x X_k + a_{Cx}\bar{X}_{t-10}$ and $\sigma \sim N(0,1)$

(d) $\bar{X}_{k-1} = \sum_{u=0}^{k-1} X_u, \bar{L}_{k-1} = \sum_{u=0}^{k-1} L_u$

6. Generate variable H_k based on inverse of the structural model $H_k = exp(\psi * \bar{X}_k)$

7. If $H_k > T^{\bar{0}}$, then observed event time $T = (t-1) + (T^{\bar{0}} - H_k - 1)exp(-\psi \bar{X}_k)$, and $Y_k = 1$ else $Y_t = 0$, set $t = t +$ increment (where increment is set to 0.1 to approximate a study in which exposure and employment status are recorded on a monthly basis) and repeat #s 6 and 7

8. Accumulate $\bar{L}_k$ and $\bar{X}_k$ to represent history variables.

9. If $H_k > T^{\bar{0}}$ and t=80, then set $Y_t = 0$ and stop (censoring by end of follow-up)

Intuitively, the simulations proceed according to the causal model proposed for healthy worker survivor bias. Methods relying on stratification to control confounding cannot remove confounding due to the time varying covariate, L_k, but methods such as g-estimation of a structural nested accelerated failure time model can remove the bias. Confounding of the relationship between cumulative exposure ($\bar{X}_k$) and the time to event outcome (T) occurs through the unblocked backdoor pathways $X_k \leftarrow V \leftarrow U \rightarrow T$ and $X \leftarrow L_k \leftarrow U \rightarrow T$ where the role of U in our example is taken the frailty indicator F. Because neither V nor L_k are causes of the outcome, the model easily allows recovery of the marginal parameter ψ, which can be interpreted as the total (net) effect of X_k on T *Rosenbaum* (1984).

We were primarily interested in the performance of the SNAFT model when no time-varying confounding (i.e. no healthy worker survivor bias) was present. The performance of SNAFT models under time varying confounding has been previously described *Young et al.* (2010). Under no time-varying confounding, we expect that both the adjusted and fully adjusted models to yield unbiased parameter estimates. This scenario can be simulated by setting $g_f = 0$ in step 6b of our data generating algorithm.

A.3.2 Simulation analyses

We fit SNAFT model models under the structural models denoted by

$$T^{\bar{0}} = m + \int_{m}^{T} exp(\psi \bar{X}_k) dk \tag{A.13}$$

We fit adjusted and unadjusted models, as in §3.1. Briefly, the exposure model for the adjusted model was fit based on the full data generating function of the data. The unadjusted model was fit to cumulative exposures conditional only on V and time.

Estimation of the SNAFT model parameters was performed by grid search parameters using the same estimation procedures described in §A.1. Due to difficulties documented by *Joffe et al.* (2012), small-sample characteristics were difficult to characterize using our estimation algorithm. Partly, this is due to the fact that the estimating function in (A.10) is not guaranteed to be smooth nor to be monotonic, as discussed for similar functions in Chapter 7 of *Kalbfleisch and Prentice* (2002). Therefore, simulations were carried out in single large samples (N=50,000) in an effort to reduce simulation bias. A preferable method would be to simulate multiple cohorts of a size that approximates the uranium miner data - however, this technique proved to be untenable for simulated samples with long follow-up. We speculate that simulating from a skewed exposure distribution often led to cohorts in which the G-function would not converge to reasonable values. For more details on why simulation in small samples is not ideal (in practice) for G-estimation, see *Joffe et al.* (2012).

Though small sample simulations proved untenable in samples that closely mimicked the miner data, we explored small sample properties using a simplified data generation algorithm. We generated the frailty variable $F = I(T^{\bar{0}} < 8)$, set the maximum time of follow-up to 6, and changed step five to be as follows:

1. V follows a Bernoulli distribution $B(p)$ where $p = Pr(V = 1) = 0.5$

2. L_k follows a Bernoulli distribution $B(p)$ where $p = Pr(L_t = 1) = expit(\boldsymbol{O}\boldsymbol{g})$, where the vectors $\boldsymbol{g} = (g_0, g_t, g_l, g_x, g_v, g_f)$ and $\boldsymbol{O} = (1, t, L_{k-1}, \bar{X}_{k-1}, V, F)$,

3. X_k follows a log-normal distribution $LN(\mu, \sigma)$ where $\mu = ln(E(X_t)) = \boldsymbol{O}^*\boldsymbol{a}$ where the vectors $\boldsymbol{a} = (a_0, a_t, a_l, a_{Cx}, a_v)$ $\boldsymbol{O}^* = (1, t, L_{k-1}, \bar{X}_{k-1}, V)$ and $\sigma \sim N(0,1)$

We emulate healthy worker survivor bias by setting values of $\boldsymbol{a}$ and $\boldsymbol{g}$ such that L_k is a time-varying confounder affected by prior exposure. Parameter values for the data generating mechanism are given in table A.3. We generated data for under time ratios of 1 and 1.14 and analyzed each sample using both a log-linear and a linear exposure model, as described above. We calculated bias as the average difference between the estimated log time ratio and the true log time ratio. Confidence intervals were estimated using the methods given in §3.1, and we estimated coverage proportion for nominal 95% confidence intervals using the sample proportion of values in which the confidence intervals contained the true time ratio. For each scenario, we fit both an adjusted and unadjusted SNAFT model, as discussed in §3.1. Some samples did not yield valid confidence intervals (i.e. the confidence region could not rule out time ratios on one side of the estimated value due to numeric difficulties), and these samples were discarded.

A.3.3 Simulation results

For large samples, under the data simulation model (Log-normal), SNAFT model results for adjusted and unadjusted models were essentially unbiased (Table A.2) for all exposure models, though the algorithm did not converge to an estimate in the unadjusted model using a log-linear exposure model. The unbiased results suggest that our unadjusted model represents a valid comparison model. For repeated samples of N=4,000, we observed that bias was essentially zero for all models in which no time-varying confounding was present, while only adjusted models were unbiased when

time-varying confounding was present. Confidence interval coverage approached nominal values for all adjusted models and for unadjusted models under no time-varying confounding when there was a non-null effect of exposure. Under no time varying confounding, coverage for unadjusted models was lower than the nominal 95% (about 70% actual coverage). This issue may have arisen due to the way in which some samples were discarded, which may have resulted in an over-representation of samples with narrow confidence regions. The nearly nominal coverage (92%) of the unadjusted model when there is a non-null exposure effect suggests that confidence interval coverage is appropriate at the nominal level.

Table A.2: Structural nested model results for simulations: large sample characteristics

Exposure model	Adjustment Set	Log time ratio	Bias*
Log-normal	Adjusted	0.075	0
	Unadjusted	NC**	NC
Log-linear	Adjusted	0.077	-0.002
	Unadjusted	0.074	0.001
Linear	Adjusted	0.074	0.001
	Unadjusted	0.074	0.001

* True log time ratio is 0.075
** Algorithm did not converge

Table A.3: Structural nested model results for simulations: small sample characteristics for N=4000, t=6, 200 samples.

Time varying confounding	Exposure model	SNAFT Model	True TR	Estimated TR[a]	Bias[b]	Coverage[c]
No[d]	Log-linear	Adjusted	1.00	0.99	-0.01	93.0
	Log-linear	Unadjusted	1.00	1.01	0.01	68.5
	Linear	Adjusted	1.00	0.99	-0.01	91.5
	Linear	Unadjusted	1.00	1.01	0.01	71.0
	Log-linear	Adjusted	1.14	1.14	0.00	95.5
	Log-linear	Unadjusted	1.14	1.13	-0.00	92.0
	Linear	Adjusted	1.14	1.14	0.00	95.5
	Linear	Unadjusted	1.14	1.13	-0.00	92.0
Yes[e]	Log-linear	Adjusted	1.00	1.01	0.01	93.0
	Log-linear	Unadjusted	1.00	0.95	-0.06	3.5
	Linear	Adjusted	1.00	1.02	0.02	92.5
	Linear	Unadjusted	1.00	0.94	-0.06	4.0
	Log-linear	Adjusted	1.14	1.13	-0.00	93.0
	Log-linear	Unadjusted	1.14	0.97	-0.16	2.0
	Linear	Adjusted	1.14	1.11	-0.03	89.5
	Linear	Unadjusted	1.14	0.96	-0.16	1.5

[b] Mean(Estimated TR)
[b] Mean(Log(true TR) - Estimated TR)
[c] 100*Proportion of confidence intervals containing true TR
[d] $\boldsymbol{g}$ = (log(1/25), log(1.140), 0, 0, log(3.5), 0), $\boldsymbol{a}$ = (log(0.2), log(0.9), 0, 0, log(3.0))
[e] $\boldsymbol{g}$ = (log(1/25), log(1.140), log(0.85), log(1/0.35), log(5.5), log(17.5)), $\boldsymbol{a}$ = (log(0.1), log(0.9), log(0.65), log(1.01), log(1.5))

A.4 Sensitivity analyses for structural nested models

A.4.1 Sensitivity analysis: exposure model used for g-estimation

A.4.1.1 Methods

Alternative exposure model To assess the sensitivity of the TR to parametric assumptions of the exposure model, we fit several SNAFT models under different assumptions about the exposure. Namely we fit the following exposure models

- Zero-inflated log-linear model (as reported in the text)
- Zero-inflated log-linear model with a log transform on time variables
- Zero-inflated linear model for exposure
- Zero-inflated log normal model for exposure, as in *Li et al.* (2011)
- Zero-inflated three parameter Gamma model for exposure
- Zero-inflated multinomial model (cumulative logit) for exposure deciles, as in *Kelley and Anderson* (2008)
- Generalized additive model for exposure

We fit the adjusted exposure model as in the text along with an "exposure adjusted only" SNAFT model in which we removed employment duration from the exposure model. The latter model provides partial adjustment for healthy worker survivor bias and may be useful for considering model fit.

A.4.1.2 Results

Similar to the model in the main text, the "zero-inflated" models are simply models with two-part likelihoods that factorize into a binary (logistic) model for any exposure

($X > 0$) and a separate Log-normal or Gamma model fit among the exposed. These have also been referred to as generalized linear models with composite links and exploded likelihoods (*Rabe-Hesketh and Skrondal* (2007)). The only models for which our SNAFT model converged to an estimate were the Zero-inflated log-linear model and the Zero-inflated linear model (Table A.4). The fully adjusted linear models yielded smaller time ratios and smaller apparent differences between exposure adjusted and fully adjusted models.

Table A.4: Comparing alternative exposure models for SNAFT model for the radon-lung cancer dose-response

			95% CI		
Exposure model	Adjustment set	Time ratio	U	L	AIC*
Zero-inflated log-linear**	Adjusted	1.166	1.152	1.174	-313202
	Exposure only	1.088	1.084	1.090	-305054
Zero-inflated log-linear	Adjusted	1.167	1.152	1.172	-313240
(log-time in model)	Exposure only	1.091	1.086	1.100	-304848
Zero inflated linear	Adjusted	1.057	1.047	1.075	-309094
	Exposure only	1.048	1.032	1.059	-305354

AIC = -2(log-likelihood - p) where p is total number of parameters estimated in the model, estimation algorithms did not converge for zero-inflated log-normal, zero-inflated gamma, or a generalized additive model, so these models are not included
**Also reported in the main text

A.4.1.3 Discussion

Because our SNAFT model model requires a model for the exposure, a model that does not accurately characterize the patterns of exposure in the CPUM data may result in bias. To see why, note that the independence function shown in 8 tests the independence between exposure and the potential outcome within strata of prior confounders. Another way to conceptualize this is to think of testing the independence of the "resid-

ual" outcome and the "residual" exposure, where the residual outcome is the part of the outcome variable that remains after subtracting (or dividing) out the part of the outcome due to exposure effects and the residual exposure is the part of the exposure variable that remains after subtracting out the part of the exposure that is predicted by prior covariates. Thus, in order for the SNAFT model to accurately characterize the exposure effects, the residual outcome and residual exposure will appear to be independent for the residual outcome at the true time ratio. However, if the exposure model is incorrect, the "residual" exposure will characterize not only the novel part of the exposure (the part of not predicted by prior covariates), but it will also reflect a poor correspondence between the model and the population function from which exposures arise.

The robustness of SNAFT model results to misspecification of a parametric exposure model is of some interest. While it is worth exploring alternative parametric or semi-parametric models for the exposure, in the CPUM there are limits to the usefulness of such an approach. Namely, exposure data are derived from records that record the date of reaching a certain cumulative exposure threshold. We created monthly exposure data using these records using a linear interpolation of the exposure rate, such that the exposure rate was assumed to be constant within each interval of time between exposure thresholds. As a consequence, model fit statistics for the exposure model are biased at best and may be misleading if used in a model selection algorithm. For example, a linear model for exposure could appear to provide a good fit to the data, but this may be a strong artifact of the way in which the data were generated. Figure A.3 shows the distribution of intervals between exposure readings in the CPUM. While a majority of the 19,260 records are more frequent than 1 year (median = 244 days), only six percent of the records are 31 days long or shorter. Thus, there is an inherent tension between capturing the variation in exposure over short intervals (to accurately capture the time-varying nature of the data) and bias in parametric modeling of exposure due

to inclusion of large amounts of linearly interpolated data.

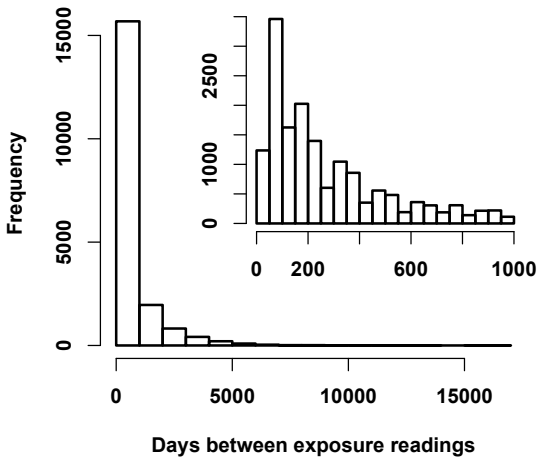

Figure A.3: Frequency of interval (days) between exposure readings in source data from Colorado Plateau Uranium miners data. Inset is histogram showing conditional distribution of intervals less than 1000 days (the largest bar in the primary figure).

Joffe et al 2012 also noted in an analysis of mortality among hemodialysis patients that some parametric exposure models did not ultimately yield estimates from a SNAFT model. This can occur either due to difficulties with an optimization algorithm or due to more fundamental issues with G-estimation. Namely, G-estimation relies on a series of nested hypothesis tests to establish independence between exposure and the potential outcome under the exposure regime "never exposed". If, for example the potential outcome is not independent of exposure at any realistic value of the time ratio, then the SNAFT model will not yield a time ratio estimate. Under the fully adjusted, zero-inflated log-linear model from the main text, the G-function (the χ^2 value testing the null hypothesis of no association between the exposure and the potential outcome under some value for the time ratio) has a clear minimum at $\psi = 0.067$ (TR=1.069, figure A4). Contrast this with the SNAFT G-function under a zero-inflated log-normal model for exposure, in which the χ^2 value never drops below 130. While the G-function clearly trends towards a minimum, the strong association between the potential outcome and

exposure suggests that the zero-inflated log-linear model for exposure does not fit the data well. Consequently, we do not report results for SNAFT models in which the exposure model was a zero-inflated log-normal model, the zero-inflated gamma model, nor a generalized additive model.

Since, to be valid, the exposure model used in G-estimation needs only to provide a valid hypothesis test of the association between the potential outcome and exposure, the linear model may seem like a useful model since it is known to be relatively robust to misspecification. However, while the results from Table A.2 show that the using a linear model for log-linearly distributed exposure data results in a relatively unbiased SNAFT parameter, the G-function may also support several different estimates of ψ. As shown in Figure A.4, which is typical of the simulation results for linear models, the G-function reaches a minimum around $\psi = 0.075$, but the confidence interval set (in which values of the χ^2 statistic is below 3.8) includes values all the way to the bottom of the grid search at $\psi = 0.03$. The G-functions fit to the CPUM data did not show this same characteristic. While existing algorithms for simulating from a SNAFT model (*Young et al.* (2010)) are useful, there is more work needed to develop more robust ways of simulating more realistic scenarios.

A.4.2 Sensitivity analysis: structural nested accelerated failure time model

A.4.2.1 Extensions of structural models

To allow for effect measure modification, we also fit a SNAFT model allowing for different time ratios per 100 WLM by duration of exposure and attained age, following the practices outlined in major publications for estimating population impact of residential radon exposures using miner data.(BEIR IV, VI, ICRP). This requires fitting a SNAFT model of the following form:

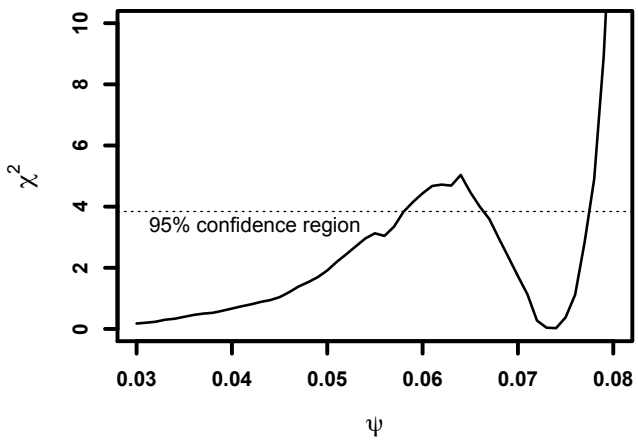

Figure A.4: G-function using a grid search algorithm for a simulated data set for 50,000 individuals. The figure shows data from the fully-adjusted linear model (Row 1 in table A.2).

$$T^{\tilde{\mathbf{0}}} = m + \int_{m}^{T} (1 + \psi_1 \bar{X}_{k-60} + \psi_2 \bar{X}_{k-60} L_k) d\,k \tag{A.14}$$

Where $\bar{X}_{k-60}$ is the cumulative exposure to radon with a 5 year (60 month) lag, ψ_1 is the log time-ratio for a 100 WLM increase in cumulative exposure in the time-varying covariate L_k, ψ_1 is the log time-ratio for a 100 WLM increase in the cumulative exposure in the referent level of the time-varying covariate and $\exp(\psi_2)$ is the factor by which the time ratio is multiplied per unit increase L_k. As in Joffe et al (2012), algorithms did not converge in models for effect measure modification and we do not report results.

A.4.2.2 Leveraging the conditional exchangeability assumption to ameliorate bias from exposure measurement error

Because radon estimates were likely much more error prone in 1950 (BEIR VI), the conditional independence function may be subject to bias when including data from this period. Thus, as reported in the text, we fit a model in which our exposure model was restricted to times after 31 December 1950. The structural model is still fit to the

entire data set, but, assuming that radon affects lung cancer with similar magnitude throughout the study period, we need only show that the potential outcome is independent of exposure in a portion of the data (Joffe et al 2010). This model somewhat reduces bias due to radon measurement error. Because this model relies on a restricted set of data, we collapsed birth cohorts into four groups in the exposure model. Compared with the TR/100 WLM (95%CI) from the adjusted log-linear SNAFT model (shown in Table A.5) of 1.088 (1.081, 1.092), the TR after fitting a model to post 1950 data was slightly higher 1.102 (1.087, 1.112). This analysis suggests that exposure misclassification may be lead to a downward bias in the Colorado Plateau uranium miner data, which agrees qualitatively with a previous report by *Stram et al.* (1999).

A.4.2.3 Latency

To assess sensitivity of our results to assumptions about latency, we fit an additional models assuming a lag of 0 years. The TR/100 WLM (95%CI) from the adjusted log-linear SNAFT model was 1.027 (1.024, 1.030) (not shown), which was slightly lower than the TR under a lag of 5 years; 1.035 (1.028, 1.039) shown in Table A.5. The g-estimation algorithm did not converge under a 0 year lag for lung cancer for the linear SNAFT model, but yielded an adjusted TR/100 WLM (95%CI) of 1.113 (1.105, 1.134) (not shown), which was slightly lower than the TR under a 5 year lag.

A.4.2.4 Non-malignant causes of death

To assess possible artifactual associations between mortality and radon exposure, we fit an additional model for non-malignant causes of death under a linear-SNAFT model, which yielded an adjusted TR/100 WLM (95%CI) of 0.987 (0.984, 0.996). The lack of positive dose response suggests that the dose responses observed between radon and lung cancer and radon and all-cause mortality is not explained by confounding by other mining related exposures, such as silica, that may have non-malignant effects.

Table A.5: Sensitivity analysis: alternative SNAFT models for quantifying the healthy worker survivor bias

		Lung Cancer				All Causes			
Model*		TR	95% CI		% diff	TR	95% CI		% diff
Log-linear	1	1.088	(1.081,	1.092)	ref	1.035	(1.028,	1.039)	ref
	2	1.044	(1.042,	1.045)	-49	1.019	(1.017,	1.023)	-44
	3	1.040	(1.039,	1.043)	-54	1.012	(1.011,	1.013)	-65
Linear excess	1	1.167	(1.152,	1.174)	ref	1.054	(1.041,	1.068)	ref
	2	1.088	(1.084,	1.094)	-47	1.027	(1.021,	1.031)	-50
	3	1.102	(1.099,	1.112)	-39	1.014	(1.013,	1.015)	-74

* Log-linear SNAFT model versus linear excess (from primary analyses) 1=Adjusted, 2=Remove employment history, 3=Unadjusted

A.5 Comparing structural nested model estimate with that from a conventional regression model

A.5.1 Methods

We also fit parametric accelerated failure time models with similar adjustment sets to the SNAFT models. We also refer to these as adjusted, and unadjusted. We assumed a generalized gamma distribution for the baseline survival distribution and transform the parameters in the model so that a TR>1 indicates a deleterious effect of exposure. All covariates used for confounder control are shown in Table 1. A comparison between the parametric models is used to quantify the magnitude of healthy worker survivor bias under the assumption that radon exposure does not affect the rate of leaving work. To facilitate comparison, we fit a log-linear SNAFT model, rather than the linear SNAFT model that was reported in the text

A.5.2 Results

Under a parametric accelerated failure time model with an assumed gamma distributed baseline survival function, the TR (95% CI) was 1.020 (1.012, 1.027) for the adjusted model and 1.037 (1.027, 1.047) for the unadjusted model. In these models, there was an 86-92% increase in the TR upon removing employment status from the model. Note that this is the opposite direction from that observed in the log-linear SNAFT models (Table A.6). Recall that the parametric models will be biased when exposure influences employment status.

A.5.3 Discussion

In a sensitivity analysis, regression models to adjust for time-varying confounding, yielded a 92% decrease in the TR in the adjusted model relative to an unadjusted model. We hypothesize that both unadjusted and adjusted parametric models are bi-

Table A.6: Comparing SNAFT model with parametric AFT model

Model		TR	95% CI		% diff
SNAFT	Fully adjusted	1.088	(1.081,	1.092)	ref
	Unadjusted	1.040	(1.039,	1.043)	-54
Parametric AFT	Fully adjusted	1.020	(1.012,	1.027)	ref
	Unadjusted	1.037	(1.027,	1.047)	86

ased because of employment status acts as a confounder and, in the miner data, cumulative exposures were associated with the hazard of leaving work. In contrast, the adjusted SNAFT model will be unbiased. Also note that the time ratio from the unadjusted SNAFT model is similar to that from the parametric model, supporting our use of the unadjusted SNAFT model as a comparison model. The TRs from Table A.6 differ from those in the primary analysis because we use a log-linear SNAFT model, which is on the same scale as the parametric model, leading to more comparable models.

A.6 Years of life lost due to occupational radon exposure during follow-up

A.6.1 Distribution of survival times

As reported in §3.1, SNAFT models yield an estimate of the expected time at death from the cause under investigation, had exposure been eliminated during follow-up. In the miner data, this corresponds to an intervention in which occupational radon exposure could be eliminated for each miner after his first health interview, such as by a mandatory respirator program. Occupational radon exposures (as shown in Figure A.5) are much higher than residential exposures which, as noted in §C.1, would be average approximately 16 WLM for an individual who lives to age 70. Thus, this potential survival time T^0 can be viewed as only a slight over-estimate of the survival time we would observe if radon exposure could be reduced to background levels through more realistic interventions, such as strong mine ventilation programs.

The distribution of the observed time of death due to lung cancer is shown (in gray) in Figure A.6. Additionally, this figure shows the distribution of the the age at death from lung cancer we would expect under an intervention to prevent exposure during follow-up. Because we would expect that this intervention would extend the lifetime, some individuals would survive past the end of follow-up in 2005 and are considered censored on that date. As noted by *Picciotto et al.* (2014), this analysis makes the assumption that an individual who is observed to die from a specific cause would also die from that cause under any hypothesized intervention on exposure. Thus, the estimate of 6,071 years of life lost among 617 lung cancer cases due to radon exposure during follow-up (Figure A.6) can be considered an upper bound because any competing causes of death could only reduce this number. For all cause mortality, our estimate of 10,118 years of life lost among the cohort (Figure A.7)) is more easily interpretable because there are no competing causes of death and loss to follow up is extremely rare in this cohort (Table 3.1).

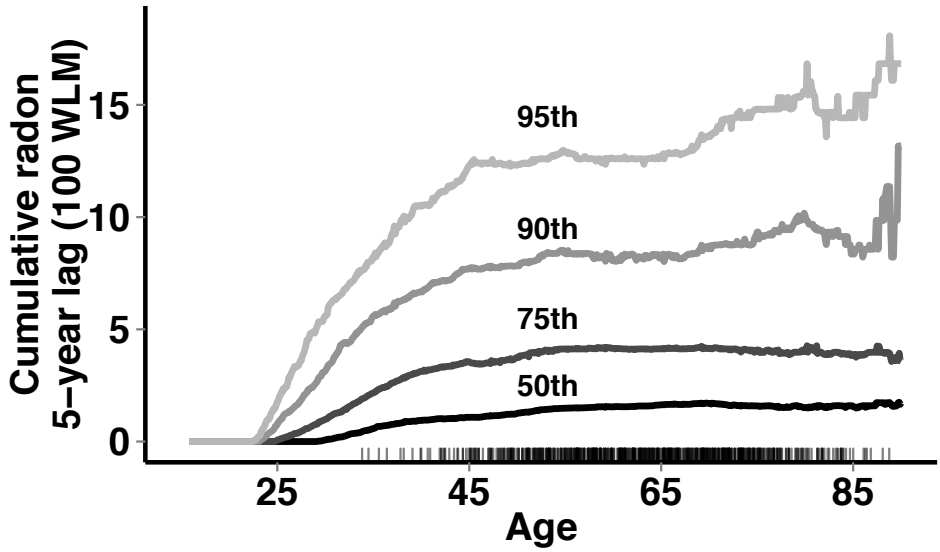

Figure A.5: Percentiles of cumulative radon exposure by age and ages of lung cancer deaths (rug plot) in the miner data.

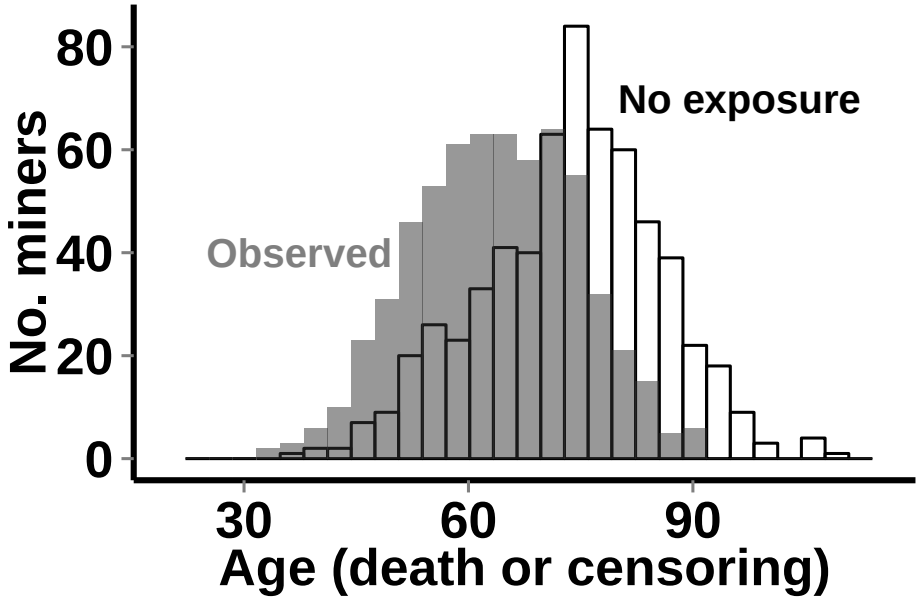

Figure A.6: Distribution of age at death from lung cancer (gray bars), and potential age at death from lung cancer or artificial censoring under no exposure (black outlined bars) in the miner data.

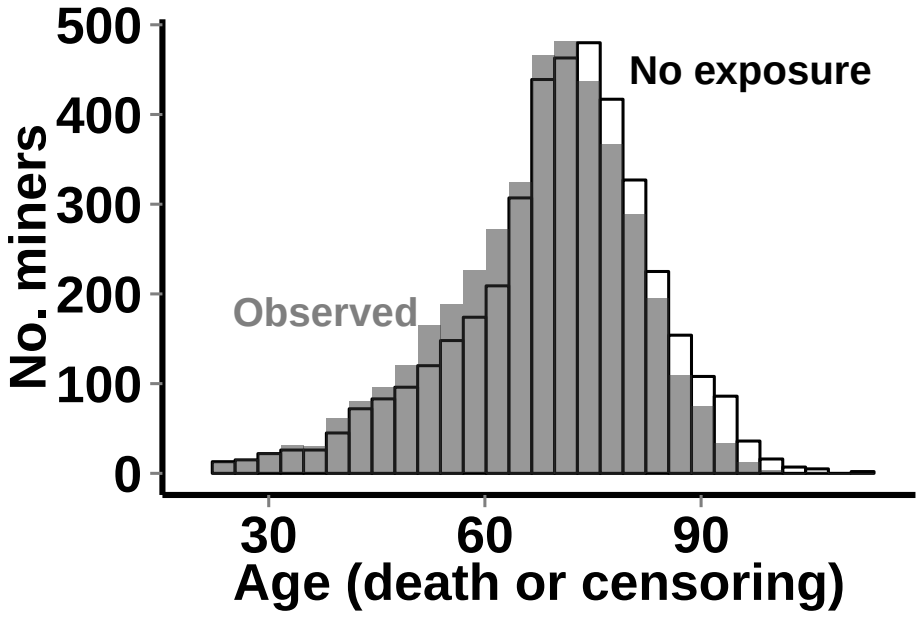

Figure A.7: Distribution of age at death from any cause (gray bars), potential age at death from any cause or artificial censoring under no exposure (black outlined bars) in the miner data.

A.6.2 Previous estimates

We identified one other study that has quantified the years of life lost due to exposure in the miners cohort (*Park et al.* (2002)). Using life-table methods, the authors calculated the excess years of potential life lost, a metric that characterizes the years lost, relative to life expectancy, in the excess cases of lung cancer in a sub-cohort of 2,721 white men hired between 1950 and 1963. These authors estimated that the excess years of potential life lost among the 267 lung cancer cases (1,271 total deaths) in the sub-cohort was 4,338 (8,431). These numbers are not directly comparable to the estimates given in the current manuscript due to differences in the cohort definition.

Another difference between these authors' analysis and the current manuscript is that the current manuscript focuses on years of life lost due to occupational radon exposure, whereas *Park et al.* (2002) estimated the excess years of life lost due to all exposures that occur in the mining environment. Thus, additional years lost due to other

mining exposures such as silica, diesel exhaust, gamma radiation, and uranium dust, may result in substantial differences between results of the current manuscript and those of *Park et al.* (2002). Finally, the estimates of these authors are likely subject to underestimation due to differences in life expectancy between the general population and miners, the latter of whom may experience, on average, lower background mortality rates. That is, the healthy hire effect (healthy hire bias) may result in an underestimate. For lung cancer, efforts were made to reduce this bias by making assumptions about the background rate of lung cancer that would be expected, had the miners not smoked nor been exposed to radon.

A.6.3 Methodologic considerations

Morfeld (2004) discusses the strong assumptions necessary to apply a causal interpretation for years of life lost in cohort analyses. The author stresses that estimates of these parameters may be biased when there are competing risks, as we discussed above. Using inverse probability of censoring weights helps to ameliorate this problem to some extent, provided that we view the estimate of years of life lost due to exposure among lung cancer cases as an upper bound. In part, the bias of the years of life lost due to exposure for a specific cause of death is rooted in an assumption inherent in survival analysis models regarding independence between disease rates of the cause of interest and the rates of censoring events, conditional on covariates. We considered competing events (for lung cancer) to be censoring events and accounted for censoring using inverse probability weights.

A.6.4 Biological considerations

An underlying assumption not noted by *Morfeld* (2004) relates to the consistency assumption noted in §C.3. That is, under a model in which we propose that exposure shifts the age at death from lung cancer, a causal interpretation may be complicated

if prognosis following lung cancer incidence differs according to age. For example, we observe in §3.1 that radon reduces the expected age at which death from lung cancer occurs. If an individual who develops incident lung cancer at age 55 would otherwise develop lung cancer at age 70, we would not expect survival following cancer initiation would be identical for the individual at both ages. This could occur due to the correlation between young age at diagnosis and tumor aggression. Alternatively, younger individuals tend to respond better to treatment than older individuals, thus increasing survival among younger cases. *Albain et al.* (1991) observed in a cohort of 2,531 extensive-stage non-small-cell lung cancer patients that age was associated with improved 1-year survival in females (34% survival for age > 70, 11% for age < 45) but not males (not reported), suggesting that survival differences by age may not be important in the miner data. By the consistency assumption, we assume that individuals who would have their lifetimes extended by eliminating exposure would have similar prognoses (had they been unexposed) as individuals who developed lung cancer at older ages.

Lung cancer is an ideal health outcome for studying mortality because of the relatively low survival of individuals with incident lung cancer. For example, *Ries et al.* (2005) report that the 5-year survival rate for lung cancer in the United States was 15.7% from 1995-2001, whereas it was only 12.5% from 1974 to 1976, and *Howlader et al.* (2013) report 5-year survival to be 17% from 2004 to 2010. Thus, this particular aspect of the consistency assumption may be weaker than studies of say, female breast cancer in which women with certain breast cancer subtypes experience 5-year survival rates as high as 95% (*O'Brien et al.* (2010)).

Few miner studies have data on incident lung cancer. However, *Lane et al.* (2010) observed that within seven strata of cumulative radon exposure, the relative rates of lung cancer incidence and lung cancer mortality were within 10% of each other. This observation suggests that results from miner studies are not subject to substantial vi-

olations from this aspect of the consistency assumption.

A.7 Left truncation in structural nested accelerated failure time models

The analyses described in §3.1 and §3.2 both utilize age as the time scale. A second approach, using time-on-study as the time scale, was not used because of the strong dependence of carcinogenic processes on age. Further, use of age as the time scale has been recommended for epidemiologic studies due to potential residual bias from covariates that might vary with age *Thiébaut and Bénichou* (2004), but that the choice should be purpose-driven *Cheung et al.* (2003). Partly, use of this time scale is motivated by the idea that two individuals of equal age are more likely to have similar hazards for cancer than are two individuals who begin follow-up at the same time but are very different ages. As shown in Figure A.8, time-on-study roughly, but not perfectly, corresponds to calendar time. Thus, another motivating factor is the lack of clear interpretation of time-on-study as a time scale. Because we can model two out of the list comprising age, period, and birth cohort, age was a clear choice of time scale because it has a clear interpretation and it is relevant to this list of time-related factors important to cancer trends.

Use of age as the time scale may be problematic if not everyone enters follow-up at the same age. This process is referred to as late entry or left truncation. If cohort members that enter the study at one age are not exchangeable with respect to the outcome under study (i.e. their hazards are identically distributed), then interpretability is compromised. This compromise occurs because we are interested in some marginal or conditional distribution of the age at lung cancer mortality, but the person-time distribution would no longer reflect our population of interest if individuals enter late, because entering into the study would depend on non-random processes. For example, if smoking miners tend to die much earlier, then the proportion of smokers in the cohort will underestimate the proportion in the target population. This exchangeabil-

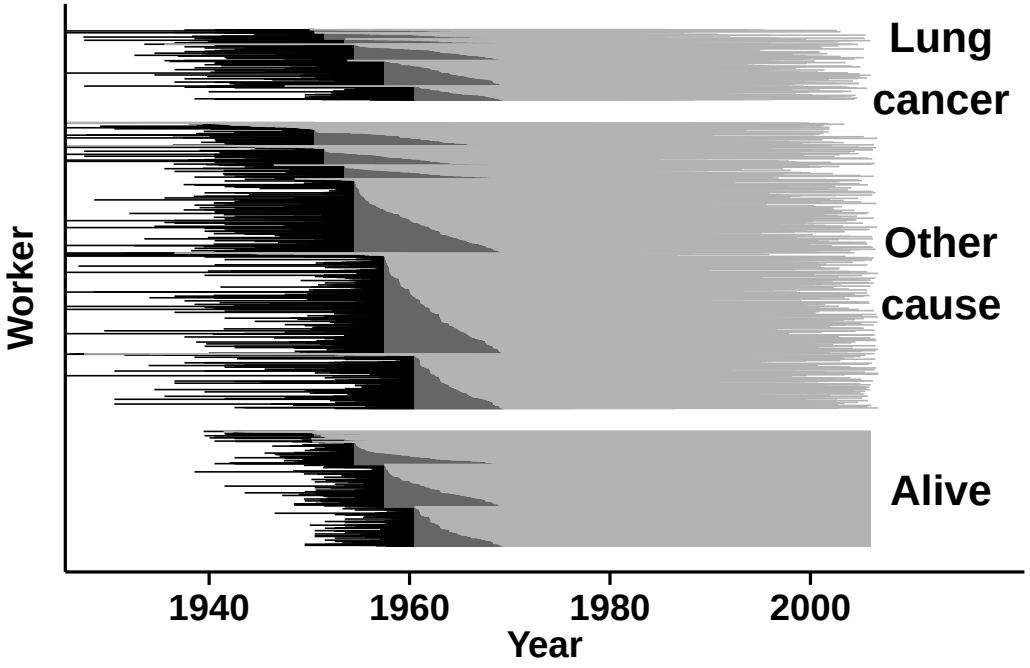

Figure A.8: Calendar time distribution of pre-follow-up employment time (black), employment time under observation (medium gray) and post-employment time (light gray) in the miner data - each horizontal line represents a single miner. The lines are grouped according to vital status at exit from the study.

ity assumption is similar to that made with respect to censoring: individuals who are in the study represent those who have left the study (or have yet to enter the study). If this assumption holds within strata of measured covariates, then late entry does not change the interpretation of our results, provided we adjust for those covariates. We could, for example, assume that smokers who are in the study are exchangeable with smokers who died before entering the study, so that the population is exchangeable within strata of the measured covariates. In this case, we would say that late entry is ignorable in strata of smoking. This can generalize to other covariates.

In the structural models presented in §3.1 and §3.2, we may be sensitive to assumptions about late entry. As shown in Figure A.9, there was variation in the age of starting the study. In the methods of §3.2, we adjusted the marginal structural model for baseline covariates. Assuming that late entry is ignorable given the baseline covariates, the marginal structural model is not subject to interpretation problems due to late entry. For the methods in §3.1, the structural nested model was marginal with respect to study covariates, and so remains subject to this problem with late entry. Because all structural nested models are subject to this caveat, we expect that the time-ratio may incur some bias. However, left truncation could not explain our observation that healthy worker survivor bias may be present in analyses with the miner data.

There are no current examples of correcting for late entry in structural nested models. We note that previous examples of structural nested models have circumvented this problem by using time-on-study as the time scale *Witteman et al.* (1998); *Hernán et al.* (2005); *Chevrier et al.* (2012); *Naimi et al.* (2014a) Analogous issues arise with right-censoring in our study, in that we assume that censoring is ignorable, given the measured covariates. Censoring is handled in §3.1 and §3.2 using inverse probability of censoring weights. These weights correct for non-ignorable censoring by applying weights to individuals in the study so that they represent themselves and any similar individuals who have been censored. A similar approach may be possible with respect

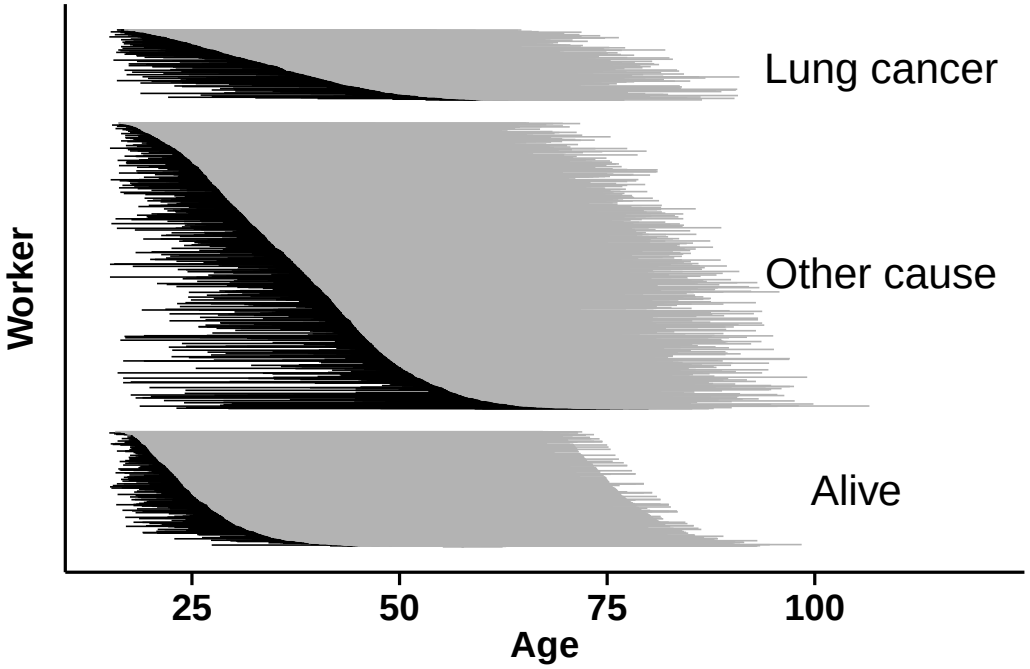

Figure A.9: Age distribution of pre-follow-up employment time (black), and time under observation (gray) - each horizontal line represents a single miner. The lines are grouped according to vital status at exit from the study.

to late entry, in which individuals on the study can be weighted in such a way that they represent both themselves and individuals who have yet to enter the study. Development of such a technique may prove valuable for both marginal structural models and structural nested models.

APPENDIX B: FURTHER DETAILS OF MARGINAL STRUCTURAL MODELS IN OCCUPATIONAL STUDIES

B.1 Why inverse probability weighted MSMs with non-positivity cannot be treated like a sparse data problem

In tabular methods, it is common to introduce what is called a continuity correction when there are empty cells in a table (*Plackett* (1964)). We continue our example from table 2 in the main text and show the application of a simple continuity correction in inverse probability weighting.

We consider the data in Table B.1, note that this is identical to the data in table 2 of the main text, with two exceptions.

1. We have added the (mean) potential outcomes $Y(00)$, $Y(10)$, $Y(01)$, $Y(11)$ to the table. These correspond to the average outcome we would observe in each stratum, had exposure been set to the value in the parentheses. For example, $Y(01)$ corresponds to the average outcome we would observe if we could set $X_1 = 0$ and $X_1 = 1$. By the consistency assumption, we note that the observed outcome for each stratum is equal to the potential outcome, given the actual exposure received (*VanderWeele* (2009); *Cole and Frangakis* (2009))

2. We have added 0.5 to each cell as a continuity correction in the calculation of the weights, so there are technically no empty cells in the weight model. To add concreteness, we refer to this added 0.5 as a "half-individual." In this case, note that the weights for the empty strata are no longer undefined, since there is a non-zero probability of falling into each stratum.

Fitting an inverse probability weighted marginal structural model using these data does not result in an unbiased estimate of the mean difference. To see this, note that, even if we supplement the weight data with 0.5 individuals in each cell, the true outcome for those individuals is unknown. We have the luxury of seeing the potential

outcomes, and, were we able to directly impute the potential outcomes, we could obtain an unbiased estimate using inverse probability weighting. Indeed, we estimate an unbiased mean difference of 18 using a marginal structural model fit using inverse probability weighting when we can correctly impute the potential outcome for the two half-individuals who are not at work but are exposed.

However, we cannot observe these outcomes. In simple tabular data, we could, in principal, impute these outcomes. Note, however, that the weights in these two strata are well over 100, implying that the results will be highly sensitive to assumptions about what those potential outcomes actually are. Also note that, if we knew the potential outcomes to begin with, there would be no need to fit a model, since we could infer the dose response from a simple comparison of potential outcomes.

As an example, note that in Table B.2 we utilize a continuity correction in both the weight estimation and the marginal structural model. Using the observed data and four possible sets of assumptions, we imputed the outcomes for the two half-individuals who were considered to be exposed off work. Our four possible assumptions based on the observed outcomes are:

No effect of employment status on the outcome The potential outcome for the half-individuals is equal to the observed outcome for those who are exposed at work

No effect of exposure on the outcome The potential outcome for the half-individuals is equal to the observed outcome for those who are exposed at work.

Max outcome The potential outcome for the half-individuals is equal to the largest observed outcome.

Min outcome The potential outcome for the half-individuals is equal to the smallest observed outcome.

The mean differences under each of these assumptions are all biased by at least 50%, and the bias is greater than that observed for the adjusted regression model in the main

text. Unfortunately, the range of these mean differences does not contain the true mean difference, so this approach is not useful for deriving bounds on the causal effect of exposure. Note that the true potential outcomes that would be observed ($Y(01) = 71$ in the second row of Table B.2 and $Y(11) = 61$ in the fifth row) are not equal to the observed outcome for any strata. Indeed, they are higher than the average observed outcome in any other stratum.

Table B.1: Continuity correction as an attempt to salvage causal effects in data subject to nonpositivity - continuity correction for weights only

X_1	L_2	X_2	$Y(00)$	$Y(10)$	$Y(01)$	$Y(11)$	Y^*	N^{**}	SW^{x**}	pN
0	0	0	35	53	53	71	35	200	0.63	201
0	0	1	35	53	53	71	?	0	151.00	†
0	1	0	11	29	29	47	11	100	1.75	281
0	1	1	11	29	29	47	29	180	0.58	281
1	0	0	25	43	43	61	43	400	0.87	401
1	0	1	25	43	43	61	?	0	101.50	†
1	1	0	1	19	19	37	19	20	3.45	81
1	1	1	1	19	19	37	37	60	0.17	81
0	0	0	35	53	53	71	35	200	0.63	201

* Unobserved in the strata $X_1 = x_1, L_2 = 0, X_2 = 1$
** N, Weight after adding 0.5 to each cell in model for weights
† Pseudo-population continues to be undefined

Table B.2: Continuity correction as an attempt to salvage causal effects in data subject to nonpositivity - continuity correction for both weight and marginal structural model

X_1	L_2	X_2	$Y(00)$	$Y(10)$	$Y(01)$	$Y(11)$	Y^*	N^{**}	SW^{x**}	pN
0	0	0	35	53	53	71	35	200.5	0.63	201
0	0	1	35	53	53	71	?	0.5	151.0	201
0	1	0	11	29	29	47	11	100.5	1.75	281
0	1	1	11	29	29	47	29	180.5	0.58	281
1	0	0	25	43	43	61	43	400.5	0.87	401
1	0	1	25	43	43	61	?	0.5	101.5	401
1	1	0	1	19	19	37	19	20.5	3.45	81
1	1	1	1	19	19	37	37	60.5	0.17	81
0	0	0	35	53	53	71	35	200.5	0.63	201

	Assumption	MD†	%Bias
1	$E[Y(X_1=1\|L_2=0]=E[Y(X_1=1\|L_2=1]$	8	-56
2	$E[Y(X_1=1\|L_2=0]=E[Y(X_1=0\|L_2=0]$	10.5	-42
3	$E[Y(X_1=1\|L_2=0]=Max(E[Y])=43$	10.5	-42
4	$E[Y(X_1=1\|L_2=0]=Min(E[Y])=11$	-2.8	-116

* Unobserved in the strata $X_1=x_1, L_2=0, X_2=1$

** N after adding 0.5 to each cell

† Mean difference from marginal structural linear model: true mean difference is 18

B.2 Estimating an MSM using the parametric G-formula

The g-formula can be defined for the expected outcome under a static intervention d. The intervention d implies setting $X_1 = X_2 = d$, in this case, so d takes on values one or zero for the interventions always- or never-exposed. Following the notation in Robins et al 2004 (cite), we define a g-formula under intervention d to be

$$f_d(Y) = \int_{-\infty}^{\infty} f(Y|L_2, d) \times f(l_2|d) d\mu(l_2) \tag{B.1}$$

In words, the marginal probability density of Y under intervention d is equal to the integrated (over values of L_2) product of the conditional probability density of Y, given $L_2 = l_2$ and d times the conditional probability that $L_2 = l_2$, given d. Because we are interested in the marginal structural model for the mean difference per unit of exposure γ_1, where $E_d(Y) = E_0(Y) + \gamma \sum_{t=1}^{2} d_t$, we are after the expectation of Y, rather than the full distribution $f_d(Y)$. $E_d(Y)$ can be estimated as

$$E_d(Y) = E(Y|L_2 = 1, d) \times Pr(L_2 = 1|d) + E(Y|L_2 = 0, d) \times Pr(L_2 = 0|d) \tag{B.2}$$

See proof in Appendix B.3. We note that, in the data in table 2 the value $E(Y|L_2 = 0, d)$ for $d = 1$ is not identified in our data, since this is the expected outcome for an individual who is always exposed, even when unemployed. The parametric g-formula uses a model for $E(Y|L_2 = 0, d)$ that extrapolates beyond the observed data and allows estimation of the quantity $E_d(Y)$.

A parametric g-formula analysis of the data using a Monte Carlo algorithm in table 2 proceeds as follows:

1. Fit a predictive logistic model for L_2: $Pr(L_2 = 1|X_1; \alpha) = expit(\alpha_0 + \alpha_1 X_1)$

2. Fit a predictive linear model for Y: $E(Y|X_1, L_2, X_2; \beta) = \beta_0 + \beta_1 X_1 + \beta_2 L_2 + \beta_3 X_2$

3. Create a "pseudo-data" dataset with a large sample size (e.g. 2,000,000), and randomize $X_1 = X_2 = 1$ (always exposed) or $X_1 = X_2 = 0$ (never exposed) for each "pseudo-individual."

4. Impute values for L_2 by taking a Bernoulli draw from the distribution $Pr(L_2 = 1|X_1 = x_1)$, where X_1 is set by the intervention

5. Impute values for Y as equal to $E(Y|X_1 = x_1, L_2 = l_2, X_2 = x_2)$, where X_1 and X_2 are set by the intervention and L_2 is predicted in step 4.

6. In the pseudo-data, fit a linear model $E(Y|X_1, X_2; \gamma) = \gamma_0 + \gamma_1(X_1 + X_2)$ the estimate of γ_1 is the g-formula standardized mean difference.

We present R code for this algorithm in appendix 3, and the pseudo-data created by this algorithm are shown in Table B.3.

As noted in the text, a marginal structural model fit with this parametric g-formula algorithm yields an unbiased mean difference of 18. To see that this is unbiased, note that, using the potential outcomes from table A1, one could simply infer the true mean difference of 18 in any stratum by taking the difference of the potential outcomes:

$$Y(11) - Y(01) = Y(11) - Y(10) = Y(01) - Y(00) = Y(10) - Y(00) = (Y(11) - (Y00))/2 = 18$$

In data, one would typically be interested in the variance of the mean difference, which is straightforward to estimate using bootstrap analysis.

Table B.3: Pseudo-data created in parametric g-formula algorithm used to estimate the mean difference per unit of cumulative exposure

X_1	L_2	X_2	$E[Y]$	
0	0	0	416741	35
0	1	0	582425	11
1	0	1	833851	61
1	1	1	166983	37
		Total	200	

B.3 Algebraic proof of the g-formula results for table 2

The mean difference of the outcome Y between interventions $d \in (0,1)$ correspond to always or never exposed, can be stated as:

$$E_1(Y) - E_0(Y) \tag{B.3}$$

The quantity $E_d(Y)$ can be expressed as

$$E_d(Y) = E(Y|L_2 = 1, d) \times Pr(L_2 = 1|d) + E(Y|L_2 = 0, d) \times Pr(L_2 = 0|d) \tag{B.4}$$

PROOF

Under the assumptions of consistency and no unmeasured confounding the g-formula for the probability distribution of Y under intervention d given in *Robins et al.* (2004)

$$f_d(y) = \int_{-\infty}^{\infty} f(y|l_2, d) \times f(l_2|d) d\mu(l_2)$$

The expectation of the outcome can be expressed as:

$$E_d(Y) = \int_{-\infty}^{\infty} f_d(Y) dy$$

If L_2 is binary

$$f_d(Y) = f(Y|L_2 = 1, d) \times Pr(L_2 = 1|d) + f(Y|L_2 = 0, d) \times Pr(L_2 = 0|d)$$

and

$$E_d(Y) = \int_{-\infty}^{\infty} [f(Y|L_2 = 0, d) \times Pr(L_2 = 0|d) + f(Y|L_2 = 1, d) \times Pr(L_2 = 1|d)] dy$$

$$= E(Y|L_2 = 0, d) \times Pr(L_2 = 0|d) + E(Y|L_2 = 1, d) \times Pr(L_2 = 1|d)$$

If the mean difference (γ) per unit of exposure X_t is defined in the following equation:

$$E_d(Y) = E_0(Y) + \gamma \sum_{t=1}^{2} d_t \tag{B.5}$$

Where $\sum_{t=1}^{2} d_t$ is 0 for the intervention never exposed and 2 for the intervention always exposed. Therefore, for the data in table 2, the marginal structural model parameter can be defined as

$$\gamma = \frac{E_d(Y) - E_0(Y)}{2} \tag{B.6}$$

Which can be estimated using a linear model for $E_d(Y)$ given the cumulative exposure under intervention d. That is, each individual can be assigned his or her expected outcome under the intervention and a regression run with that expected outcome as the dependent variable and the cumulative exposure under the intervention as the sole dependent variable.

B.4 R code to perform the g-formula analysis given in the main text

```
#carry out g-formula using simple example from table 2
#Step 0: read in data
dat <- structure(list(X1 = c(0, 0, 0, 0, 1, 1, 1, 1),
                      L2 = c(0, 0, 1, 1, 0, 0, 1, 1),
                      X2 = c(0, 1, 0, 1, 0, 1, 0, 1),
                      N = c(200, 0, 100, 180, 400, 0, 20, 60),
                      Y = c(35, NA, 11, 29, 43, NA, 19, 37)),
                      .Names = c("X1", "L2", "X2", "N", "Y"),
                      row.names = c(NA, -8L), class = "data.frame")

#Step 1: predictive model for L2, given X1
#pr(L|X1)
ml <- glm(L2 ~ X1, weight=N, data=dat, family="binomial")

#Step 2: predictive model for Y, given X1, X2, L2
#f(Y|L, X1, X2)
my <- glm(Y ~ L2 + X1 + X2, weight=N, data=dat)

#Step 3: generate randomized exposures (always vs. never)
set.seed(1983430)
mc.size = 2e6 #Monte Carlo dataset size
X1 <- X2 <- rbinom(mc.size, 1, 0.5)
mc.dat <- data.frame(X1, X2)

#Steps 4,5: impute values for L2 and Y based on predictive
#distributions from step 2
```

```
#always exposed
mc.dat$L2 <- rbinom(mc.size, 1, predict(ml, type="response",
                    newdata=mc.dat))
mc.dat$Y <- predict(my, type="response", newdata=mc.dat)

#Step 6: Fit a linear model for the cumulative exposure
#model for g-formula standardized mean difference
glm(Y ~ I(X1+X2), data=mc.dat)

# Coefficients:
# (Intercept)   I(X1 + X2)
      # 21.01        18.00
# The estimate is unbiased
```

APPENDIX C: FURTHER ISSUES IN CAUSAL INFERENCE AND RADON

C.1 Notes on units of ionizing radiation used in this manuscript

We report multiple units of radiation in the manuscript. The following provides a rough conversion for working level months and picoCuries/L and provides some context for the levels of exposure discussed in this manuscript versus that which we observe in modern residences. These conversions contain many assumptions and should be taken to provide conversion with a large margin of error - nonetheless, they may be useful to understand the order of magnitude of each exposure measure.

- Units of radiation quantities reported in the current text

  working level months WLM - any combination of exposure rate in working levels [130,000 MeV of potential α energy per liter of air] and employment time that leads to an average exposure of 1 working level over 170 hours.

  bequerels - Bq - nuclear disintegrations per second in a volume of air (see *ICRU* (2011b))

- The US Environmental Protection Agency "action level" for radon remediation (taking action to reduce indoor radon levels) is 4 picoCuries/Liter (pCi/L *USEPA* (2003)).

- 1 pCi/L corresponds to 37 bequerels per cubic meter Bq/m^3, so the action level corresponds to approximately 150 Bq/m^3 (*USEPA* (2003)).

- The average exposure over 70 years at average domestic radon levels in the United States is equal to 16 working level months (*Lubin et al.* (1997)).

- Radon exposure occurs at approximately 1.8 working level months/year (WLM/y) in a home with a radon concentration of 370 Bq/m^3 (*Lubin et al.* (1997)).

- A miner exposed to 25 working level months is assumed to have approximately the same exposure as an individual living 25 years in a home with a 370 Bq/m^3

radon concentration (*Lubin and Boice* (1997)).

Cumulative radon daughter exposures are measured in working level months (WLM), a unit devised originally for occupational applications. Exposure is proportional to concentration (WL) and time, with exposure to 1 WL for 170 h being defined as 1 WLM. To convert from residential exposures expressed in pCi/L, the BEIR VI committee assumed that the fraction of time spent indoors is 70%. It follows that an indoor radon concentration of 1 pCi/L would on average result in an exposure of 0.144 WLM/y = (1 pCi/L) [(0.7)(0.004) WL/(pCi/L)] (51.6 WLM/WL-y)

-*USEPA* (2003) p 17

C.2 Critical issues for inference in studies radon and lung cancer

Cohort studies, such as the miner studies discussed above, are potentially rich sources of information regarding relationships between environmental and occupational agents and health outcomes. Their longitudinal nature allows for nuanced questions regarding the evolving web of complex relationships between various health indicators. Radon exposure occurs over the entire life course, so miner studies with long exposure history and extended follow-up for disease status are invaluable at answering such questions. Healthy worker survivor bias is one concern about how a dynamic interplay between factors in the occupational environment and employment may bias any parameter that attempts a simultaneous analysis of exposures that occur across the lifespan. This dynamism also allows researchers to answer other questions relating to the time-varying factors that may influence the strength of associations between radon and disease. Any attempt to address healthy worker survivor bias should be cognizant of such issues. In both the miner literature and the wider literature on the effects of ionizing radiation on disease, lines of question have addressed modification by smoking, attained age, age at exposure, exposure rate, and time since exposure. Below, we address some of the literature on these factors as well as concerns about possible confounding by smoking and the impact of measurement error in miner studies.

C.2.1 Radon and smoking

Tobacco smoking is strongly related to rates of lung cancer in humans. Effect measure modification and confounding of the radon-lung cancer association are both of interest. With respect to effect measure modification, it is often noted that smoking may be an important modifier in both miner studies (*NRC* (1999); *UNSCEAR* (2008)) and in pooled residential studies (*Darby et al.* (2005); *Krewski et al.* (2005)). When (*NRC* (1999)) analyzed animal data from COGEMA (France), they observed synergism

(higher than expected joint effects, given the marginal effects) when radon dose was followed by a dose of cigarette smoke, but no synergism was found if the order was reversed. The data from Colorado Plateau uranium miners includes data on smoking taken from initial surveys and subsequent follow-up (*Archer et al.* (1976); *Schubauer-Berigan et al.* (2009)). In the Colorado Plateau uranium miners, (*Schubauer-Berigan et al.* (2009)) found a super-additive, sub-multiplicative interaction between radon and smoking status at the most recent survey. *Tomásek* (2011) found similar interaction, intermediate between additive and multiplicative, reflecting lower baseline rates of cancer among non-smokers (ERR/WLM among smokers = 0.011, among non-smokers = 0.044). *Leuraud et al.* (2011) fit a mixing model in which a parameter which ranged from 0 (fully additive) to 1 (fully multiplicative) indicating the model of best fit for the smoking-radon interaction, and they observed the lowest deviance model at a value of 0.5 for the mixing parameter, indicating a super-additive, sub-multiplicative interaction.

While exposure to tobacco smoke may modify the effect of radon on lung cancer, the evidence suggest that confounding by smoking in miner studies is small. *Schnelzer et al.* (2010) found in a case-cohort study of German miners in which smoking history was taken from retrospective interviews and medical records abstraction that the unadjusted ERR/ WLM was 0.0025, whereas after adjustment for smoking the ERR/ WLM was 0.0023, and a similarly small change in estimate was observed in a French case-control study among miners (unadjusted ERR/WLM = 0.011, smoking adjusted ERR/WLM = 0.007) (*Amabile et al.* (2009)). *Bijwaard et al.* (2011) investigated the confounding effects of smoking in an analysis in which smoking rates from a case-control study were projected onto a cohort of German miners. The investigators applied mechanistic models to the data. Similar to the animal models of COGEMA, the investigators observed that smoking mainly acts in later rate-limiting steps of cancer development, but that including smoking in the model did not change risk projections. In a

joint analysis of three case-control studies of miners (French, Czech, and German), (*Leuraud et al.* (2011)) observed an unadjusted ERR/WLM of 0.010 and a smoking adjusted ERR/WLM of 0.008, suggesting a minimal effect of smoking on the radon-lung cancer dose response. Identifying confounding by smoking using the change in estimate approach (which is the approach informally used here) is complicated by non-collapsibility of the ERR as well as the apparent effect measure modification by smoking. Because smoking is a strong risk factor for lung cancer, one could reasonably expect non-collapsibility of the ERR over smoking strata. While no correlation between smoking intensity and radon exposure rates were given, *Hornung et al.* (1998) found that including cumulative smoking in lung cancer models did not appreciably change the relative rate for lung cancer per unit of radon exposure in the Colorado Plateau Uranium miners. There is little *a priori* evidence to suspect that smoking levels are associated with radon exposure, and observational studies of miners do have evidence that such an association is strong enough to result in appreciable bias.

C.2.2 Attained age

Because radon exposures occur over the life course, it is difficult to disentangle time-related effects of exposure. For example, estimating the effects of exposure received at young ages (age at exposure) will be complicated by considering by time since exposure and attained age because the latter two will by correlated in individuals exposed at similar times. One approach is to look at these factors individually, or mutually adjusted for each other. For example *NRC* (1999) reported models in which the association between cumulative radon exposure and lung cancer was stratified over levels of attained age and windows of time since exposure. The authors estimated a decreasing relative risk with attained age in 10 year categories of age over age 55. This approach may be useful when focused on risk prediction when the specific source of the variation may be less important than accurately estimating stratum spe-

cific rates or risks. A similar approach was used in a recent analysis of the Eldorado (Canadian) uranium miner cohort, in which the relative rates of lung cancer mortality per 100 WLM were modified by a factor of 1 (ref), 1.62, 0.82, 0.19 for attained ages $< 55, 55-64, 65-74,$ and 75+ (*Lane et al.* (2010)). This pattern held for lung cancer incidence, as well. Similarly, *Leuraud et al.* (2011) noted monotonically decreasing RRs across age groups $< 55, 55-65, > 65$ while not adjusting for other time related factors, a pattern that is potentially explainable as either an artifact of the effect measure - lung cancer incidence increases with age, so relative risks will decrease for linear dose-response relationships - or as a fading of the effectiveness of dose over time since exposure, a pattern observed in latency analysis in the CPUM (*Langholz et al.* (1999)). Alternatively, a fading effect over attained age is also consistent with an accelerated life/accelerated failure-time model in which radon exposure influences lung cancer rates by accelerating the time to lung cancer incidence, rather than strictly by causing cases that would otherwise not occur.

C.2.3 Age at exposure

As noted below, different models exist for the modifying effects of age at exposure. In a joint study of Czech and French uranium miners (*Tomásek et al.* (2008)), for example estimated that the ERR for lung cancer (adjusted for cumulative exposure) was 0.52 per decade of age at median exposure (centered at 30), indicating that exposed person-time in older ages incurred a lower relative risk of lung cancer than exposed person-time at younger ages. In a similar model among German uranium miners (*Walsh et al.* (2010b)) observed a RR per decade of 0.70. Effects of age on exposure have not been explored in recent updates of the Colorado Plateau uranium miners (*Hornung et al.* (1998); *Schubauer-Berigan et al.* (2009)). *Leuraud et al.* (2011) observed an RR per decade of 0.70 in a joint case-control study of three countries. The observations of lower relative risks for lung cancer from radon exposure at older ages agrees with other

findings from individuals exposed to ionizing radiation in a cohort of atomic bomb survivors (*Preston et al.* (2007)) and several cohorts of nuclear workers (*Richardson et al.* (2001)).

C.2.4 Exposure rate

At an average of 800 WLM per five years of exposure, the Colorado Plateau uranium miners exposure concentrations are high relative to the other miner cohorts included in the BEIR VI report and relative to residential exposures which average 14 WLM over a lifetime (*NRC* (1999)). Exposure rates were higher in earlier periods than in later periods in this cohort (Figure C.1). Extrapolation of risk parameters from miner studies to other populations exposed at lower dose rates is hampered by potential dose-rate effects, namely what has been termed the "inverse-dose-rate-effect." (*Darby and Doll* (1990)) showed that the summary ERR/WLM for each miner cohort in the BEIR VI report was smaller for cohorts with a higher dose rate. (*NRC* (1999)) truncated person time to lower cumulative exposure levels due to non-linearity of the effect after including higher doses, which all came from the Colorado Plateau uranium miners. At relatively high levels of radiation, cell killing effects can reduce the effectiveness of a given dose of radiation, but at lower doses, microdosimetric factors such as interaction between radiation exposures and the cell cycle may influence the dose rate effect (*Brenner and Hall* (1990); *Brenner et al.* (1993)).

In the Colorado Plateau uranium miners (*Hornung et al.* (1998)) observed a negative coefficient for average exposure rate of -0.013 (RR = 0.98) per WL in a model that also adjusted for time since last exposure, and smoking. Initial model building indicated a possible non-linear effect of dose, but when exposure rate was included in the model, the non-linear term did not improve fit of the model. In a joint analysis of Czech and French miners, a cohort with much lower dose-rate than that in the Colorado Plateau uranium miners, stratification of the ERR for cumulative measured (rather

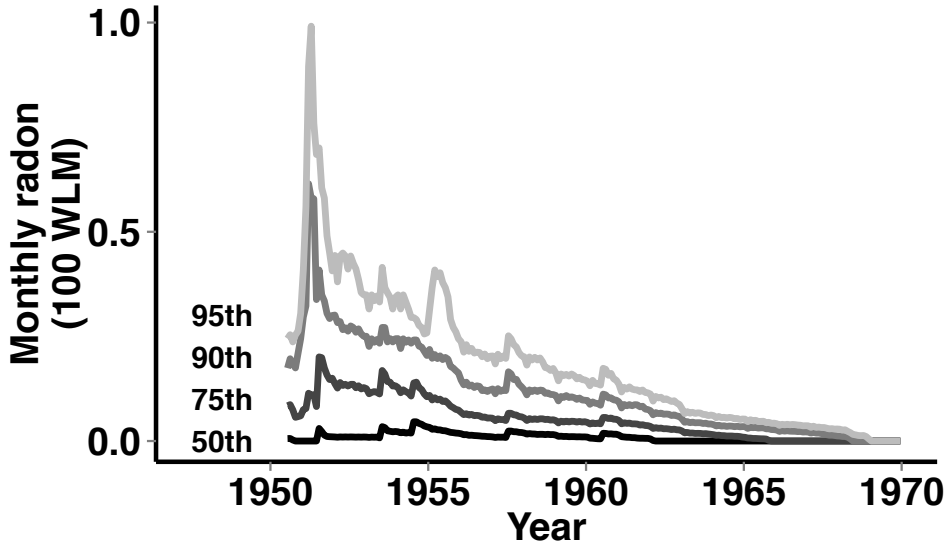

Figure C.1: Percentiles of monthly radon exposure by year in the Colorado Plateau uranium miners.

than estimated) exposure by categories of exposure rate below 4 WL did not improve model fit in a model that included time since median exposure and age at median exposure. The ERR/WLMs in each category of exposure rate ($< 0.5, .5-1, 1-2, 2, > 4$ WL) were 0.060, 0.018, 0.042, 0.040, 0.032, which did not differ substantially from the summary ERR/WLM of 0.042 in a model controlling for age at median exposure and time since median exposure. In the largest uranium miners cohort in Germany, (*Walsh et al.* (2010b)) observed a monotonic decrease in the ERR/WLM over categories of increasing exposure rate, with categories $< 0.5, .5-1, 1-3, 3-5, 5-15, 15+$ (ERR/100 WLM = 6.14, 3.56, 2.76, 2.35, 1.98, ref), consistent with the patterns observed in the BEIR VI report (*NRC* (1999)).

C.2.5 Latency and time since exposure

A final time related aspect of the radon lung cancer association involves the time period over which an increment of exposure to radon affects lung cancer rates, and the time at the peak, rate of increasing effects to reach that peak, and the rate of de-

cline of the effect after the peak. Investigators often specify *a priori* a lag of five years when studying the association between radon exposure and lung cancer. Under this assumption, exposures incurred within the previous five years of an observation are ignored while exposures after 5 years are weighted equally. However, unless the interval between exposure and outcome in a cohort is exactly equal for all members, this leads to misclassification of the etiologically relevant exposure. *Langholz et al.* (1999) demonstrated the utility of a simple bivariate lag function that weights exposures using a triangle-shaped kernel, such that exposures proximal and distal from to the observed period are down-weighted, but in some intermediate period exposures are given a full weight of one, corresponding to the observed peak effect of exposure. This method offers a simple parametric way to address latency in occupational cohorts, and is easily applied to other other models (*Richardson* (2009b)).

(*Hornung et al.* (1998)) modeled time since last exposure as a linear term in an ERR model including exposure rate and smoking and observed a negative slope (RR=0.96) per year of time since last exposure. *Tomásek* (2011) observed a monotonic decrease in the ERR/WLM over categories of increasing time since exposure (starting with 5-19 years) across estimates of a mixture model used to assess the nature of the effect measure modification between smoking and radon. Among German uranium miners (*Walsh et al.* (2010a)) observed an RR of 0.5 per decade of time since last exposure (as a log linear term), indicating agreement with other cohorts and with (*NRC* (1999)).

C.2.6 Misclassification and measurement error

Measurement error (misclassification, in categorical variables), or differences between the measured and actual values, can affect both the validity and the the precision of effect estimates when considering error in measuring the exposure and the outcome (*Rothman et al.* (2008)), chapter 9. Measurement error of covariates can also reduce the ability to control for confounding and selection bias (*Marshall et al.* (1999);

Fewell et al. (2007)).

Outcome misclassification Estimation of effects of exposures incurred at younger ages on chronic conditions that occur most often in older ages requires long-term follow-up of workers. In the Colorado Plateau uranium miners cohort, this has been achieved through data-linkage with the National Death Index and the Social Security Administration death files (*Roscoe* (1997); *Schubauer-Berigan et al.* (2009)) or death certificate data before the advent of the National Death Index (*Whittemore and McMillan* (1983)). Since, in the case of the Colorado Plateau uranium miners, the highest exposures occurred in the 1950s and 1960s (*NRC* (1999)), early loss to follow up would be likely to impact the precision of the estimates, but it is not likely to impact estimates for all-cause mortality, since loss to follow up in the Colorado Plateau uranium miners cohort was 0.3% before 1979 (*Schubauer-Berigan et al.* (2009)). *Cowper et al.* (2002) summarized the accuracy of both the National Death Index and the Social Security administration using studies that could validate the sensitivity and specificity of the searches using previously validated deaths National Death Index sensitivity estimates varied from 87% to 98%, while specificity varied from 92% to 100%.

Potential misattribution of causes of death are of greater concern than unidentified deaths. Cause of death in the Colorado Plateau uranium miners has been determined by review of death certificates, following determination of mortality and follow up with death registries for the location of the death certificate (*Sathiakumar et al.* (1998)). *Schubauer-Berigan et al.* (2009) reported that, cause of death could not be determined for fewer than 1% of miners, indicating that nearly all of the miners have some attributed cause. However, the authors observed an excess of deaths from silicosis and postulated that, while miners may have been exposed to silica dust, it was also likely that radon-related disease was misclassified as silicosis for some of the miners. *Doria-Rose and Marcus* (2009) reported a sensitivity of around 89% and a specificity

of 99% for accurate coding of lung cancer as the cause of death among participants in the Mayo Lung Project in Minnesota from 1971-1983. Since radon was a suspected lung carcinogen, sensitivity may be better in the Colorado Plateau uranium miners cohort. In the French uranium miner cohort, (*Laurier et al.* (2004)) found that identified errors in the cause of death impacted SMRs, but did not substantially bias dose-response estimates. The impact of outcome misclassification likely depends not the nature of the disease of interest. Diseases with poor prognosis and high case-fatality rates may be less likely misidentified than other diseases. Additionally, mortality data for diseases with low survival also allow one to make inferences about incidence, whereas effect measures for other diseases may be affected by factors such as treatment or prognostic factors related to exposure. For white, US males in 2007, the age-standardized incidence rate of lung cancer was 84.9 per 100,000 men, while the mortality rate was 69.8 (*Kohler et al.* (2011)). By contrasting the incidence/mortality ratio for lung cancer ($84.9/69.8 = 1.2$) to that for all cancers combined ($552.5/225.4 = 2.5$), the use of lung cancer as an outcome in a study utilizing death certificate data is less prone to bias due to outcome misclassification than would be a study based on use of all cancer mortality (and, by extension, many other possible outcomes of interest).

Exposure measurement error Exposure measurement is of concern in the Colorado Plateau uranium miners due not only to attenuation of dose-response estimates due to errors in personal exposure estimates, but also due to variation across time in the quality of radon measurements. In the Colorado Plateau uranium miners, exposure measurement error has been documented to occur due to both the stochastic processes of measurement devices as well as to systematic processes. As shown in figure C.2, exposure estimates from the Colorado Plateau uranium miners are based solely on area measurements, area estimations (average of several measurements in a single area), extrapolations (taking estimates from previous or subsequent years), and, in

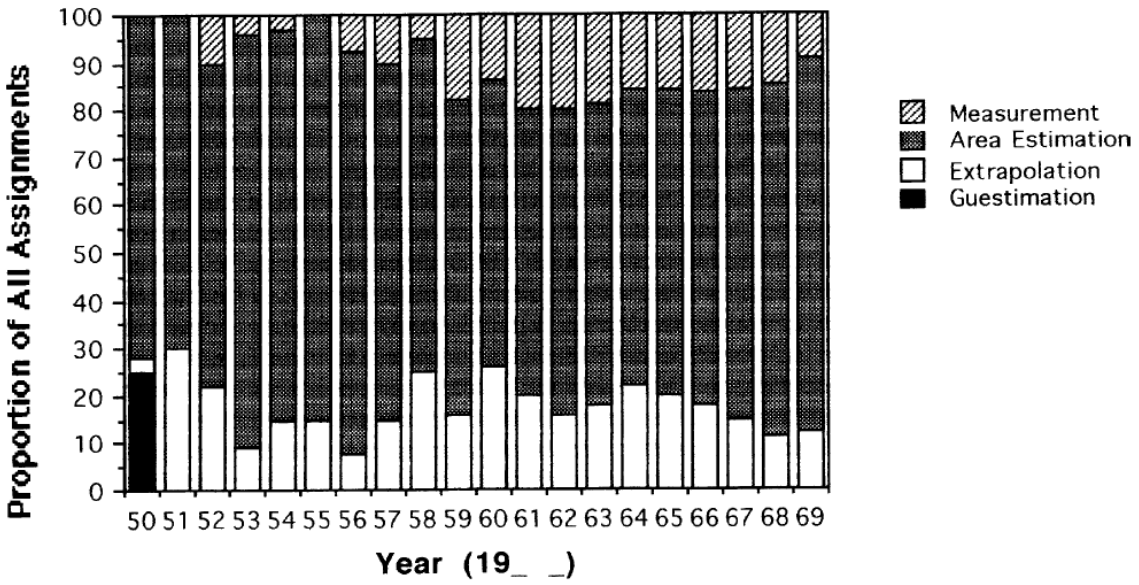

Figure C.2: Sources of exposure assignment in the Colorado Plateau uranium miners, 1950-1969. Source (*NRC* (1999)), Annex E

1950, what (*NRC* (1999)) terms as "Guestimation" where mine experts judged the probable radon levels using physical characteristics of the mine and measurements performed at later dates. Most of the exposure estimation was performed by taking area estimates based on nearby (if possible) measurements. Typically, the type of measurement error in which group averages are assigned to an individual result in increased uncertainty, but little attenuation of dose response (*Armstrong* (1998)). However, the machines that measured radon concentrations were subject to errors as well, which would tend to attenuate dose responses. Further, the quality of measurements varied across time, as did exposure levels, so the quality of measurement is associated with cumulative exposures, thus complicating interpretation of dose-response estimated for cumulative exposure when all mine data are used. In a simulation based on a French uranium miner cohort (*Allodji et al.* (2012)) proposed that measurement error contributes to a 60% attenuation of the ERR per 100 WLM. Within the actual cohort, (*Vacquier et al.* (2009)) observed an order of magnitude increase in dose-response estimates for exposed person time beginning in 1956, when personal doses were dramatically decreased and the number and quality of exposure measurements dramatically

increased. In a joint analysis of the French and Czech cohorts, (*Tomásek et al.* (2008)) estimated an ERR/WLM that was 80-85% higher in the person-years with high quality personal dosimetery versus the ERR/WLM in person-years with estimated exposures based on area measures. In the WISMUT (German) cohort of uranium miners, (*Walsh et al.* (2010b)) observed in an ERR model adjusted for age-at- and time-since-median exposure an ERR/WLM of 0.45 in person-time with measured exposures, and 0.34 in person-time with estimated exposures. Because exposure rates tend to correlate with measurement quality, the influence of measurement error and dose-rate-effects is difficult to disentangle.

Some exposure measurements in the Colorado Plateau uranium miners are subject to systematic errors, as well. *Schiager* (1989) noted an early practice of taking area level radon progeny measurements during only the summer months, when air circulation may be lowest due to reduced convection. *Lundin Jr et al.* (1971) also noted that, because measurements made after 1960 were made for control purposes, they systematically over-estimated the exposures after that time. An attempt at dose correction by *Stram et al.* (1999) indicates that the dose-response may be under-estimated in this cohort, though their analysis was limited by its assumption that there were no systematic measurement errors of the type cited above. Many exposure estimates made after 1960 were excluded due to suspicion that the exposure measurements (made by the mining companies for regulatory control) may have been "reduced to avoid regulatory action" (*NRC* (1999)), p. 308, but the excluded estimates are lost to time and the impact of measurement culling is unknown.

Covariate misclassification: employment status, age, time aspects, smoking Employment status, whether or not a miner was employed in a particular year, is an essential factor for controlling bias by the methods proposed in this manuscript. A general rule of thumb is that measurement errors in covariates considered for confound-

ing (consider employment status to be a member of the set of covariates to control confounding of the radon-lung cancer association) affects the ability to control confounding by that covariate, leading to residual confounding (*Armstrong* (1998)). Work history in the Colorado Plateau uranium miners is assigned by self-report from questionnaires given to the miners by the US Public Health Service as well as follow-up interviews, which ceased in 1969. Misclassification of work time has not previously been explored in the Colorado Plateau uranium miners, thus, the extent to which residual confounding may bias estimates is unknown.

In the statistical methods used in the current manuscript (structural nested accelerated failure time models and marginal structural models), employment status is considered as a confounder because it affects subsequent exposure (i.e. unemployed individuals are, by definition, unexposed) and is likely associated with the age at lung cancer mortality either through an unmeasured common cause (e.g. frailty - sick individuals are more likely to terminate) or because it is a causal risk factor for lung cancer mortality (e.g. through changes in behavioral risk factors following termination). Employment status changes over time, and it is thus termed a time-varying confounder. Because job termination is a memorable event, it is unlikely that an event such as retirement or termination from a mining job is substantially mis-measured. However, intermittent work stoppages, which might occur if a particular uranium deposit was exhausted and a miner spent some time finding a job with a new mining company, may be subject to more substantial error. Further, if these intermittent stoppages are related to health status (e.g. if illness caused a temporary break from work), then adjustment for employment status will result in some residual confounding.

It is also possible that miners moved from a mine that was included in the Colorado Plateau uranium miners to one that was not included, in which case termination date (and exposures) would be under-estimated.

Perhaps a more fundamental issue of consideration of employment status as a con-

founder is that, while it is generally considered as a binary indicator of current employment in the industry of interest, the effect of employment status on mortality is likely heterogeneous. For example, a worker that terminates in one industry to move to another likely has a different health status from one who terminates in an industry and does not seek new employment due to chronic illness. Additionally, occupational exposures likely influence subsequent employment status in a heterogeneous way, as well. Given a similar time-since-hire, individuals who terminate due to early stage respiratory disease are likely to have higher occupational radon exposures than individuals who leave uranium mining for other industries. Unfortunately, individual reasons for employment status changes are not recorded in Colorado Plateau uranium miners data.

Measurement error and misclassification in other covariates reduces the ability to control for confounding and biases stratum specific measures for misclassified effect measure modifiers of the radon-lung cancer association. Because the timing of employment is self-reported, there is potential for misclassification of the factors discussed in §C.2. Smoking habits were recorded in surveys in the Colorado Plateau uranium miners into the 1970s, but recent smoking follow-up has been less complete, necessitating interpolation and extrapolation of smoking habits (*Roscoe* (1997); *Schubauer-Berigan et al.* (2009)). Assessment of effect measure modification by strata of ever/never smoking may be less prone to bias than assessment on categories of smoking intensity, but doing so also restricts the validity of effect measures, since someone who smoked for a year as a young person is likely at much lower risk of lung cancer than a lifetime smoker.

C.2.7 Summary

Miner studies are useful for addressing questions about time-related issues in the radon-lung cancer association, due to the extensive measurement history and long-

term follow-up. Many of these time related aspects function as effect measure modifiers that can be explored using the methods proposed in Chapter II.

C.3 Sufficient assumptions for a causal interpretation of estimated associations

To employ a causal interpretation, all methods all rely on the identifying assumptions of conditional exchangeability, consistency, non-interference, and positivity. These assumptions can be expressed as follows, where T^z is the failure-time we would observe under the intervention of setting $Z = z$, as in a clinical trial where z could be "take the placebo."

These are often referred to as identifiability assumptions, which are reviewed in detail by *Daniel et al.* (2013).

Conditional exchangeability $X_k \perp T^{x_k} | \bar{\boldsymbol{X}}_{\boldsymbol{k-1}}, \bar{\boldsymbol{L}}_{\boldsymbol{k}}, \boldsymbol{V}$ where $A \perp B|C$ is interpreted as "A is independent of B, conditional on C." The outcome we would observe under exposure at time k being set to $X_k = x_k$ is independent of the actual exposure at time k, given the past values of the exposure and covariates (cf *Greenland and Robins* (2009))

Treatment version irrelevance $T^{\bar{\boldsymbol{x}}_k} = T|\bar{\boldsymbol{X}}_{\boldsymbol{k}} = \bar{\boldsymbol{x}}_{\boldsymbol{k}}$. The outcomes we would observe if we set $\bar{\boldsymbol{X}}_{\boldsymbol{k}} = \bar{\boldsymbol{x}}_{\boldsymbol{k}}$ in some intervention is the same set of outcomes we would observe if individuals were observed to have $\bar{\boldsymbol{X}}_{\boldsymbol{k}} = \bar{\boldsymbol{x}}_{\boldsymbol{k}}$ (cf *Cole and Frangakis* (2009); *VanderWeele* (2009))

Non-interference $T^{\bar{\boldsymbol{x}}_{ki}\bar{\boldsymbol{x}}_{kj}} = T^{\bar{\boldsymbol{x}}_{ki}}$, for all units $i \neq j$. The potential outcome of an individual is independent of the hypothetical treatment regime of any other unit (cf *Rosenbaum* (2007))

Positivity $f(X_k|\bar{\boldsymbol{X}}_{k-1}, \bar{\boldsymbol{L}}_{\boldsymbol{k}}, \boldsymbol{V}) > 0$ with probability 1 if $f(\bar{\boldsymbol{X}}_{k-1}, \bar{\boldsymbol{L}}_{\boldsymbol{k}}, \boldsymbol{V}) > 0$. The probability (density) of exposure taking on any value is non-zero for every observed strata of prior covariates and exposure (cf appendix A.1 of *Hernán and Robins* (2006), *Westreich and Cole* (2010))

REFERENCES

Albain, K. S., J. Crowley, M. LeBlanc, and R. Livingston (1991), Survival determinants in extensive-stage non-small-cell lung cancer: the southwest oncology group experience, *J Clin Oncol*, *9*(9), 1618–1626.

Albert, A., and J. Anderson (1984), On the existence of maximum likelihood estimates in logistic regression models, *Biometrika*, *71*(1), 1.

Allodji, R. S., K. Leuraud, A. C. M. Thiébaut, S. Henry, D. Laurier, and J. Bénichou (2012), 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*, doi:10.1007/s00411-012-0403-3.

Amabile, J.-C., K. Leuraud, B. Vacquier, S. Caër-Lorho, A. Acker, and D. Laurier (2009), Multifactorial study of the risk of lung cancer among french uranium miners: radon, smoking and silicosis, *Health Phys*, *97*(6), 613–21, doi:10.1097/01.HP.0000363842.62922.58.

Applebaum, K. M., E. J. Malloy, and E. A. Eisen (2007), Reducing healthy worker survivor bias by restricting date of hire in a cohort study of vermont granite workers, *J Occup Environ Med*, *64*(10), 681–7, doi:10.1136/oem.2006.031369.

Applebaum, K. M., E. J. Malloy, and E. A. Eisen (2011), Left truncation, susceptibility, and bias in occupational cohort studies, *Epidemiology*, *22*(4), 599–606, doi:10.1097/EDE.0b013e31821d0879.

Archer, V. (1962), Hazards to health in uranium mining and milling, *J Occup Environ Med*, *4*(2), 55.

Archer, V. E., J. D. Gillam, and J. K. Wagoner (1976), Respiratory disease mortality among uranium miners, *Ann N Y Acad Sci*, *271*, 280–93.

Armstrong, B. G. (1998), Effect of measurement error on epidemiological studies of environmental and occupational exposures, *J Occup Environ Med*, *55*(10), 651–6.

Arrighi, H., and I. Hertz-Picciotto (1994), The evolving concept of the healthy worker survivor effect, *Epidemiology*, *5*(2), 189–96.

Arrighi, H., and I. Hertz-Picciotto (1996), Controlling the healthy worker survivor effect: an example of arsenic exposure and respiratory cancer, *J Occup Environ Med*, *53*(7), 455–62.

Arrighi, H. M., and I. Hertz-Picciotto (1993), Definitions, sources, magnitude, effect modifiers, and strategies of reduction of the healthy worker effect, *J Occup Med*, *35*(9), 890–2.

Arrighi, H. M., and I. Hertz-Picciotto (1995), Controlling for time-since-hire in occupational studies using internal comparisons and cumulative exposure, *Epidemiology*, *6*(4), 415–8.

Bergdahl, I. A., H. Jonsson, K. Eriksson, L. Damber, and B. Järvholm (2010), Lung cancer and exposure to quartz and diesel exhaust in swedish iron ore miners with concurrent exposure to radon, *J Occup Environ Med*, *67*(8), 513–8, doi:10.1136/oem.2009.047456.

Bijwaard, H., F. Dekkers, and T. van Dillen (2011), Modelling lung cancer due to radon and smoking in wismut miners: preliminary results, *Radiat Prot Dosimetry*, *143*(2-4), 380–3, doi:10.1093/rpd/ncq469.

Björ, O., H. Jonsson, L. Damber, J. Wahlström, and T. Nilsson (2013), Reduced mortality rates in a cohort of long-term underground iron-ore miners, *Am J Ind Med*, *56*(5), 531–40, doi:10.1002/ajim.22168.

Brenner, A. V., et al. (2011), I-131 dose response for incident thyroid cancers in ukraine related to the chornobyl accident, *Environ Health Perspect*, *119*(7), 933–9, doi:10.1289/ehp.1002674.

Brenner, D., and E. Hall (1990), The inverse dose-rate effect for oncogenic transformation by neutrons and charged particles: a plausible interpretation consistent with published data, *Int J Radiat Biol*, *58*(5), 745–758.

Brenner, D., E. Hall, G. Randers-Pehrson, and R. Miller (1993), Mechanistic considerations on the dose-rate/let dependence of oncogenic transformation by ionizing radiations, *Radiat Res*, *133*(3), 365–369.

Brenner, D. J. (2009), Extrapolating radiation-induced cancer risks from low doses to very low doses, *Health Phys*, *97*(5), 505–9, doi:10.1097/HP.0b013e3181ad7f04.

Brenner, D. J., and R. K. Sachs (2006), Estimating radiation-induced cancer risks at very low doses: rationale for using a linear no-threshold approach, *Radiat Environ Biophys*, *44*(4), 253–6, doi:10.1007/s00411-006-0029-4.

Breslow, N., and N. Day (1987), Statistical methods in cancer research. volume ii-the design and analysis of cohort studies. iarc scientific publication number 82, *Lyon, France: International Agency for Research on Cancer*.

Buckley, J. P., A. P. Keil, L. J. McGrath, and J. K. Edwards (2014), Evolving methods for inference in the presence of healthy worker survivor bias, *In preparation*.

Cain, L. E., and S. R. Cole (2009), Inverse probability-of-censoring weights for the correction of time-varying noncompliance in the effect of randomized highly active antiretroviral therapy on incident aids or death, *Stat Med*, *28*(12), 1725–38, doi:10.1002/sim.3585.

Cardis, E., et al. (2007), The 15-country collaborative study of cancer risk among radiation workers in the nuclear industry: estimates of radiation-related cancer risks, *Radiat Res*, *167*(4), 396–416.

Chen, J., L. Bergman, R. Falcomer, and J. Whyte (2014), Results of simultaneous radon and thoron measurements in 33 metropolitan areas of canada., *Radiat Prot Dosimetry*, doi:10.1093/rpd/ncu141.

Cheung, Y. B., F. Gao, and K. S. Khoo (2003), Age at diagnosis and the choice of survival analysis methods in cancer epidemiology, *Journal of clinical epidemiology*, *56*(1), 38–43.

Chevrier, J., S. Picciotto, and E. A. Eisen (2012), 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*, *23*(2), 212–9, doi:10.1097/EDE.0b013e318245fc06.

Chiba, Y., K. Azuma, and J. Okumura (2009), Marginal structural models for estimating effect modification, *Ann Epidemiol*, *19*(5), 298–303, doi:10.1016/j.annepidem.2009.01.025.

Christiani, D. C., T. T. Ye, S. Zhang, D. H. Wegman, E. A. Eisen, L. A. Ryan, S. A. Olenchock, L. Pothier, and H. L. Dai (1999), Cotton dust and endotoxin exposure and long-term decline in lung function: results of a longitudinal study., *Am J Ind Med*, *35*(4), 321–331.

Cole, S. R., and H. Chu (2005), Effect of acyclovir on herpetic ocular recurrence using a structural nested model, *Contemp Clin Trials*, *26*(3), 300–10, doi:10.1016/j.cct.2005.01.009.

Cole, S. R., and C. E. Frangakis (2009), The consistency statement in causal inference: a definition or an assumption?, *Epidemiology*, *20*(1), 3–5, doi:10.1097/EDE.0b013e31818ef366.

Cole, S. R., and M. A. Hernán (2002), Fallibility in estimating direct effects, *Int J Epidemiol*, *31*(1), 163–5.

Cole, S. R., and M. A. Hernán (2008), Constructing inverse probability weights for marginal structural models, *Am J Epidemiol*, *168*(6), 656–64, doi:10.1093/aje/kwn164.

Cole, S. R., and M. G. Hudgens (2012), Re: "marginal structural models for case-cohort study designs to estimate the association of antiretroviral therapy initiation with incident aids or death", *Am J Ind Epidemiol*, *175*(7), 732, doi:10.1093/aje/kws170.

Cole, S. R., M. A. Hernán, J. B. Margolick, M. H. Cohen, and J. M. Robins (2005), Marginal structural models for estimating the effect of highly active antiretroviral therapy initiation on cd4 cell count, *Am J Epidemiol*, *162*(5), 471–8, doi:10.1093/aje/kwi216.

Cole, S. R., M. A. Hernan, K. Anastos, B. D. Jamieson, and J. M. Robins (2007), Determining the effect of highly active antiretroviral therapy on changes in human immunodeficiency virus type 1 rna viral load using a marginal structural left-censored mean model., *Am J Epidemiol*, *166*(2), 219–227.

Cole, S. R., R. W. Platt, E. F. Schisterman, H. Chu, D. Westreich, D. Richardson, and C. Poole (2010), Illustrating bias due to conditioning on a collider, *Int J Epidemiol*, *39*(2), 417–20, doi:10.1093/ije/dyp334.

Cole, S. R., D. B. Richardson, H. Chu, and A. I. Naimi (2013), Analysis of occupational asbestos exposure and lung cancer mortality using the g formula, *Am J Epidemiol*, *177*(9), 989–996, doi:10.1093/aje/kws343.

Cowper, D. C., J. D. Kubal, C. Maynard, and D. M. Hynes (2002), A primer and comparative review of major us mortality databases, *Ann Epidemiol*, *12*(7), 462–8.

Daniel, R. M., M. G. Kenward, S. N. Cousens, and B. L. De Stavola (2011), Using causal diagrams to guide analysis in missing data problems, *Stat Methods Med Res*, doi:10.1177/0962280210394469.

Daniel, R. M., S. N. Cousens, B. L. De Stavola, M. G. Kenward, and J. A. C. Sterne (2013), Methods for dealing with time-dependent confounding, *Stat Med*, *32*(9), 1584–618, doi:10.1002/sim.5686.

Darby, S., et al. (2005), Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 european case-control studies, *Brit Med J*, *330*(7485), 223–226, doi:10.1136/bmj.38308.477650.63.

Darby, S. C., and R. Doll (1990), Radiation and exposure rate, *Nature*, *344*(6269), 824, doi:10.1038/344824a0.

Darby, S. C., et al. (1995), Radon and cancers other than lung cancer in underground miners: a collaborative analysis of 11 studies, *J Natl Cancer Inst*, *87*(5), 378–84.

Doria-Rose, V. P., and P. M. Marcus (2009), Death certificates provide an adequate source of cause of death information when evaluating lung cancer mortality: an example from the mayo lung project, *Lung Cancer*, *63*(2), 295–300, doi:10.1016/j.lungcan.2008.05.019.

Dumas, O., N. Le Moual, V. Siroux, D. Heederik, J. Garcia-Aymerich, R. Varraso, F. Kauffmann, and X. Basagaña (2013), Work related asthma. a causal analysis controlling the healthy worker effect, *J Occup Environ Med*, *70*(9), 603–10, doi:10.1136/oemed-2013-101362.

Edwards, J., L. McGrath, J. Buckley, M. Schubauer-Berigan, S. Cole, and D. Richardson (2014), Occupational radon exposure and lung cancer mortality: estimating intervention effects using the parametric g-formula, *Epidemiology*, p. In press.

Fewell, Z., G. Davey Smith, and J. A. C. Sterne (2007), The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study, *Am J Epidemiol*, *166*(6), 646–55, doi:10.1093/aje/kwm165.

Flanders, W. D., V. M. Cárdenas, and H. Austin (1993), Confounding by time since hire in internal comparisons of cumulative exposure in occupational cohort studies, *Epidemiology*, *4*(4), 336–41.

Fornalski, K. W., and L. Dobrzynski (2011), Pooled bayesian analysis of twenty eight studies on radon induced lung cancers, *Health Phys*, *101*(3).

Fox, A., and P. Collier (1977), Mortality experience of workers exposed to vinyl chloride monomer in the manufacture of polyvinyl chloride in great britain., *Brit J Ind Med*, *34*(1), 1–10.

Fox, A. J., and P. F. Collier (1976), Low mortality rates in industrial cohort studies due to selection for work and survival in the industry, *Br J Prev Soc Med*, *30*(4), 225–30.

Garshick, E., F. Laden, J. E. Hart, B. Rosner, M. E. Davis, E. A. Eisen, and T. J. Smith (2008), Lung cancer and vehicle exhaust in trucking industry workers, *Environ Health Perspect*, *116*(10), 1327–32, doi:10.1289/ehp.11293.

Garshick, E., F. Laden, J. E. Hart, M. E. Davis, E. A. Eisen, and T. J. Smith (2012), Lung cancer and elemental carbon exposure in trucking industry workers., *Environ Health Perspect*, *120*(9), 1301–1306, doi:10.1289/ehp.1204989.

Gierl, S., O. Meisenberg, P. Feistenauer, and J. Tschiersch (2014), Thoron and thoron progeny measurements in german clay houses, *Radiat Prot Dosimetry*, doi:10.1093/rpd/ncu076.

Gilbert, E., and S. Marks (1979), An analysis of the mortality of workers in a nuclear facility, *Radiat Res*, *79*(1), 122–148.

Gilbert, E. S. (1982), Some confounding factors in the study of mortality and occupational exposures, *Am J Epidemiol*, *116*(1), 177–88.

Greenland, S. (2003), Quantifying biases in causal models: classical confounding vs collider-stratification bias., *Epidemiology*, *14*(3), 300–306.

Greenland, S., and J. Pearl (2008), Causal diagrams, *In S. Boslaugh and L.A. McNutt (Eds.) Encyclopedia of epidemiology*, *1*, 149–156.

Greenland, S., and J. M. Robins (2009), Identifiability, exchangeability and confounding revisited, *Epidemiol Perspect Innov*, *6*(1), 4, doi:10.1186/1742-5573-6-4.

Greenland, S., J. Pearl, and J. Robins (1999), Causal diagrams for epidemiologic research, *Epidemiology*, *10*(1), 37–48.

Hauptmann, M., K. Berhane, B. Langholz, and J. Lubin (2001), Using splines to analyse latency in the colorado plateau uranium miners cohort, *J Epidemiol Biostat*, *6*(6), 417–24.

Hernán, M. A., and J. M. Robins (2006), Estimating causal effects from epidemiological data, *J Epidemiol Commun H*, *60*(7), 578–586, doi:10.1136/jech.2004.029496.

Hernán, M. A., B. A. Brumback, and J. M. Robins (2000), Marginal structural models to estimate the causal effect of zidovudine on the survival of hiv-positive men., *Epidemiology*, *11*(5), 561–570.

Hernán, M. A., B. A. Brumback, et al. (2001), Marginal structural models to estimate the joint causal effect of nonrandomized treatments, *J Am Stat Assoc*, *96*, 440–448.

Hernán, M. A., S. R. Cole, J. Margolick, M. Cohen, and J. M. Robins (2005), Structural accelerated failure time models for survival analysis in studies with time-varying treatments, *Pharmacoepidem Dr S*, *14*(7), 477–491, doi:10.1002/pds.1064.

Hertz-Picciotto, I., H. M. Arrighi, and S. W. Hu (2000), Does arsenic exposure increase the risk for circulatory disease?, *Am J Epidemiol*, *151*(2), 174–81.

Holaday, D., D. Rushing, R. Coleman, P. Woolrich, and H. Kusnetz (1957), Control of radon and daughters in uranium mines and calculations on biologic effects, *Tech. rep.*, U.S. Department of Health, Education, and Welfare, Public Health Service, Bureau of State Services, Division of Special Health Services, Occupational Health Program.

Hornung, R., J. Deddens, and R. Roscoe (1998), Modifiers of lung cancer risk in uranium miners from the colorado plateau, *Health physics*, *74*(1), 12–21.

Hornung, R. W., and T. J. Meinhardt (1987), Quantitative risk assessment of lung cancer in u.s. uranium miners, *Health Phys*, *52*(4), 417–30.

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

Horvitz, D., and D. Thompson (1952), A generalization of sampling without replacement from a finite universe, *J Am Stat Assoc*, *47*(260), 663–685.

Howe, G. R., A. M. Chiarelli, and J. P. Lindsay (1988), Components and modifiers of the healthy worker effect: evidence from three occupational cohorts and implications for industrial compensation, *Am J Epidemiol*, *128*(6), 1364–75.

Howlader, N., et al. (2013), Seer cancer statistics review, 1975–2011.

IARC (1988), *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Volume 43: Man-made Mineral Fibres and Radon*, World Health Organization.

IARC (2001), *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Volume 78 : Ionizing Radiation, Part 2 : Some Internally Deposited Radionuclides*, World Health Organization.

IARC (2012), *IARC Monographs On The Evaluation Of Carcinogenic Risks To Humans, Volume 100D: A Review of Human Carcinogens: Radiation*, no. 100-D in IARC Monographs On The Evaluation Of Carcinogenic Risks To Humans, 1–341 pp., WHO International Agency for Research on Cancer Monograph Working Group.

ICRP (1993), Icrp publication 65: Protection against radon-222 at home and at work, *Ann ICRP*, *23*(2).

ICRP (2010), Icrp publication 115: Lung cancer risk from radon and progeny and statement on radon, *Ann ICRP*, *40*(1), 1–64, doi:10.1016/j.icrp.2011.08.011.

ICRU (2011a), 5. dosimetry, *Journal of the ICRU*, *11*(1), 23–28, doi:10.1093/jicru/ndr003.

ICRU (2011b), Report 85, *Journal of the ICRU*, *11*(1), NP, doi:10.1093/jicru/ndr012.

ICRU (2011c), 2. general considerations, *Journal of the ICRU*, *11*(1), 7–9, doi:10.1093/jicru/ndr002.

Jacobi, W. (1993), The history of the radon problem in mines and homes, *Annals of the ICRP*, *23*(2), 39–45.

Jiránek, M., and V. Kačmařikova (2014), Dealing with the increased radon concentration in thermally retrofitted buildings, *Radiat Prot Dosimetry*, doi:10.1093/rpd/ncu104.

Joffe, M. M. (2001), Administrative and artificial censoring in censored regression models, *Stat Med*, *20*(15), 2287–304, doi:10.1002/sim.850.

Joffe, M. M. (2012), Structural nested models, g-estimation, and the healthy worker effect: the promise (mostly unrealized) and the pitfalls, *Epidemiology*, *23*(2), 220–2, doi:10.1097/EDE.0b013e318245f798.

Joffe, M. M., D. R. Hoover, L. P. Jacobson, L. Kingsley, J. S. Chmiel, B. R. Visscher, and J. M. Robins (1998), Estimating the effect of zidovudine on kaposi's sarcoma from observational data using a rank preserving structural failure-time model, *Stat Med*, *17*(10), 1073–102.

Joffe, M. M., W. P. Yang, and H. I. Feldman (2010), Selective ignorability assumptions in causal inference, *Int J Biostat*, *6*(2), Article 11.

Joffe, M. M., W. P. Yang, and H. Feldman (2012), G-estimation and artificial censoring: problems, challenges, and applications, *Biometrics*, *68*(1), 275–286, doi:10.1111/j.1541-0420.2011.01656.x.

Jonsson, H., I. A. Bergdahl, G. Akerblom, K. Eriksson, K. Andersson, L. Kågström, B. Järvholm, and L. Damber (2010), Lung cancer risk and radon exposure in a cohort of iron ore miners in malmberget, sweden, *J Occup Environ Med*, *67*(8), 519–25, doi:10.1136/oem.2009.047449.

Kalbfleisch, J., and R. Prentice (2002), *The Statistical Analysis of Failure Time Data (Wiley Series in Probability and Statistics)*, 2 ed., Wiley-Interscience.

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

Keil, A. P., J. K. Edwards, D. B. Richardson, A. I. Naimi, and S. R. Cole (2014a), The parametric g-formula for time-to-event data: Intuition and a worked example, *Epidemiology*, doi:10.1097/EDE.0000000000000160.

Keil, A. P., D. B. Richardson, S. R. Cole, M. Troester, and M. G. Hudgens (2014b), Healthy-worker survivor bias in the colorado plateau uranium miners cohort, *In preparation*.

Kelley, M. E., and S. J. Anderson (2008), Zero inflation in ordinal data: incorporating susceptibility to response through the use of a mixture model, *Stat Med*, *27*(18), 3674–88, doi:10.1002/sim.3267.

Kohler, B. A., et al. (2011), Annual report to the nation on the status of cancer, 1975-2007, featuring tumors of the brain and other nervous system, *J Natl Cancer Inst*, *103*(9), 714–36, doi:10.1093/jnci/djr077.

Korhonen, P. A., N. M. Laird, and J. Palmgren (1999), Correcting for non-compliance in randomized trials: an application to the atbc study, *Stat Med*, *18*(21), 2879–97.

Kreuzer, M., B. Grosche, M. Schnelzer, A. Tschense, F. Dufey, and L. Walsh (2010), Radon and risk of death from cancer and cardiovascular diseases in the german uranium miners cohort study: follow-up 1946-2003, *Radiat Environ Biophys*, *49*(2), 177–85, doi:10.1007/s00411-009-0249-5.

Kreuzer, M., K. Straif, J. W. Marsh, F. Dufey, B. Grosche, D. Nosske, and M. Sogl (2012), Occupational dust and radiation exposure and mortality from stomach cancer among german uranium miners, 1946-2003, *J Occup Environ Med*, *69*(3), 217–23, doi:10.1136/oemed-2011-100051.

Krewski, D., et al. (2005), Residential radon and risk of lung cancer: a combined analysis of 7 north american case-control studies, *Epidemiology*, *16*(2), 137–145.

Krewski, D., et al. (2006), A combined analysis of north american case-control studies of residential radon and lung cancer, *J Tox Env Health, Part A*, *69*(7-8), 533–597.

Lane, R. S. D., S. E. Frost, G. R. Howe, and L. B. Zablotska (2010), Mortality (1950-1999) and cancer incidence (1969-1999) in the cohort of eldorado uranium workers, *Radiat Res*, *174*(6), 773–85, doi:10.1667/RR2237.1.

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

Laurier, D., M. Tirmarche, N. Mitton, M. Valenty, P. Richard, S. Poveda, J.-M. Gelas, and B. Quesne (2004), An update of cancer mortality among the french cohort of uranium miners: extended follow-up and new source of data for causes of death, *Eur J Epidemiol*, *19*(2), 139–46.

Le Moual, N., F. Kauffmann, E. A. Eisen, and S. M. Kennedy (2008), The healthy worker effect in asthma: work may cause asthma, but asthma may also influence work, *Am J Respir Crit Care Med*, *177*(1), 4–10, doi:10.1164/rccm.200703-415PP.

Leuraud, K., M. Schnelzer, L. Tomásek, N. Hunter, M. Tirmarche, B. Grosche, M. Kreuzer, and D. Laurier (2011), Radon, smoking and lung cancer risk: results of a joint analysis of three european case-control studies among uranium miners, *Radiat Res*, *176*(3), 375–87.

Li, N., D. A. Elashoff, W. A. Robbins, and L. Xun (2011), A hierarchical zero-inflated log-normal model for skewed responses, *Stat Methods Med Res*, *20*(3), 175–89, doi: 10.1177/0962280208097372.

Lubin, J. (1994), *Radon and lung cancer risk: a joint analysis of 11 underground miners studies*, US Dept. of Health and Human Services, Public Health Service, National Institutes of Health.

Lubin, J. H. (1988), Models for the analysis of radon-exposed populations, *Yale J Biol Med*, *61*(3), 195–214.

Lubin, J. H. (2003), Studies of radon and lung cancer in north america and china, *Radiat Prot Dosimetry*, *104*(4), 315–9.

Lubin, J. H., and J. D. Boice, Jr (1997), Lung cancer risk from residential radon: meta-analysis of eight epidemiologic studies, *J Natl Cancer Inst*, *89*(1), 49–57.

Lubin, J. H., Z. Liang, Z. Hrubec, G. Pershagen, J. B. Schoenberg, W. J. Blot, J. B. Klotz, Z. Y. Xu, and J. D. Boice, Jr (1994), Radon exposure in residences and lung cancer among women: combined analysis of three studies, *Cancer Causes Control*, *5*(2), 114–28.

Lubin, J. H., et al. (1995a), Lung cancer in radon-exposed miners and estimation of risk from indoor exposure, *J Natl Cancer Inst*, *87*(11), 817–27.

Lubin, J. H., et al. (1995b), Radon-exposed underground miners and inverse dose-rate (protraction enhancement) effects, *Health Phys*, *69*(4), 494–500.

Lubin, J. H., et al. (1997), Estimating lung cancer mortality from residential radon using data for low exposures of miners, *Radiat Res*, *147*(2), 126–34.

Lundin Jr, F., J. Lloyd, E. Smith, V. Archer, and D. Holaday (1969), Mortality of uranium miners in relation to radiation exposure, hard-rock mining and cigarette smoking-1950 through september 1967, *Health Phys*, *16*(5), 571.

Lundin Jr, F., J. Wagoner, and V. Archer (1971), *Radon daughter exposure and respiratory cancer quantitative and temporal aspects*, National Institute for Occupational Safety and Health. Joint monograph no. 1, U.S. Public Health Service.

Marshall, J. R., J. L. Hastrup, and J. S. Ross (1999), Mismeasurement and the resonance of strong confounders: correlated errors, *Am J Epidemiol*, *150*(1), 88–96.

McNamee, R. (2003), Confounding and confounders, *J Occup Environ Med*, *60*(3), 227–34.

Morfeld, P. (2004), Years of life lost due to exposure: Causal concepts and empirical shortcomings, *Epidemiol Perspect Innov*, *1*(1), 5, doi:10.1186/1742-5573-1-5.

Naimi, A. I., S. R. Cole, D. J. Westreich, and D. B. Richardson (2011), A comparison of methods to estimate the hazard ratio under conditions of time-varying confounding and nonpositivity, *Epidemiology*, *22*(5), 718–723, doi:10.1097/EDE.0b013e31822549e8.

Naimi, A. I., S. R. Cole, M. G. Hudgens, M. A. Brookhart, and D. B. Richardson (2013), Assessing the component associations of the healthy worker survivor bias: occupational asbestos exposure and lung cancer mortality, *Ann Epidemiol*, *23*(6), 334–341.

Naimi, A. I., S. R. Cole, M. G. Hudgens, and D. B. Richardson (2014a), 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*, *25*(2), 246–54, doi:10.1097/EDE.0000000000000045.

Naimi, A. I., E. E. M. Moodie, N. Auger, and J. S. Kaufman (2014b), Constructing inverse probability weights for continuous exposures: a comparison of methods, *Epidemiology*, *25*(2), 292–9, doi:10.1097/EDE.0000000000000053.

Neophytou, A. M., S. Picciotto, J. E. Hart, E. Garshick, E. A. Eisen, and F. Laden (2014), A structural approach to address the healthy-worker survivor effect in occupational cohorts: an application in the trucking industry cohort, *J Occup Environ Med*, doi:10.1136/oemed-2013-102017.

NRC (1988), *Health Risks Of Radon And Other Internally Deposited Alpha-Emitters: BEIR IV*, Washington, DC: National Academy Press.

NRC (1998), *Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII Phase I*, Washington, DC: National Academy Press.

NRC (1999), *Health Effects of Exposure to Radon: BEIR VI*, Washington, DC: National Academy Press.

O'Brien, K. M., S. R. Cole, C.-K. Tse, C. M. Perou, L. A. Carey, W. D. Foulkes, L. G. Dressler, J. Geradts, and R. C. Millikan (2010), Intrinsic breast tumor subtypes, race, and long-term survival in the carolina breast cancer study, *Clin Cancer Res*, *16*(24), 6100–10, doi:10.1158/1078-0432.CCR-10-1533.

Park, R. M., A. J. Bailer, L. T. Stayner, W. Halperin, and S. J. Gilbert (2002), An alternate characterization of hazard in occupational epidemiology: years of life lost per years worked, *Am J Ind Med*, *42*(1), 1–10, doi:10.1002/ajim.10082.

Pearce, N. (1992), Time-related confounders and intermediate variables, *Epidemiology*, *3*(4), 279–81.

Pearce, N., H. Checkoway, and D. Kriebel (2007), Bias in occupational epidemiology studies, *J Occup Environ Med*, *64*(8), 562–8, doi:10.1136/oem.2006.026690.

Pearl, J. (1995), Causal diagrams for empirical research, *Biometrika*, *82*(4), 669–688.

Petersen, M. L., S. G. Deeks, J. N. Martin, and M. J. van der Laan (2007), History-adjusted marginal structural models for estimating time-varying effect modification, *Am J Epidemiol*, *166*(9), 985–93, doi:10.1093/aje/kwm232.

Petersen, M. L., K. E. Porter, S. Gruber, Y. Wang, and M. J. van der Laan (2012), Diagnosing and responding to violations in the positivity assumption, *Stat Methods Med Res*, *21*(1), 31–54, doi:10.1177/0962280210386207.

Picciotto, S., D. M. Brown, J. Chevrier, and E. A. Eisen (2013), Healthy worker survivor bias: implications of truncating follow-up at employment termination, *J Occup Environ Med*, *70*, 736–742, doi:10.1136/oemed-2012-101332.

Picciotto, S., J. Chevrier, J. Balmes, and E. A. Eisen (2014), Hypothetical interventions to limit metalworking fluid exposures and their effects on copd mortality: G-estimation within a public health framework, *Epidemiology*, *25*(3), 436–43, doi:10.1097/EDE.0000000000000082.

Plackett, R. (1964), The continuity correction in 2×2 tables, *Biometrika*, *51*(3-4), 327.

Polley, E. C., and M. J. van der Laan (2010), Super learner in prediction, *Tech. Rep. 222*, Division of Biostatistics, University of California, Berkeley.

Preston, D., Y. Shimizu, D. Pierce, A. Suyama, and K. Mabuchi (2003), Studies of mortality of atomic bomb survivors. report 13: Solid cancer and noncancer disease mortality: 1950-1997, *Radiation Research*, *160*(4), 381–407.

Preston, D., E. Ron, S. Tokuoka, S. Funamoto, N. Nishi, M. Soda, K. Mabuchi, and K. Kodama (2007), Solid cancer incidence in atomic bomb survivors: 1958-1998, *Radiat Res*, *168*(1), 1–64.

Price, P. N., A. Nero, K. Revzan, M. Apte, A. Gelman, and W. J. Boscardin (2011), http://eande.lbl.gov/iep/high-radon/hr.html (last checked 6/2014).

Rabe-Hesketh, S., and A. Skrondal (2007), Multilevel and latent variable modeling with composite links and exploded likelihoods, *Psychometrika*, *72*(2), 123–140.

Rage, E., B. Vacquier, E. Blanchardon, R. S. Allodji, J. W. Marsh, S. Caër-Lorho, A. Acker, and D. Laurier (2012), Risk of lung cancer mortality in relation to lung doses among french uranium miners: Follow-up 1956-1999, *Radiat Res*, *177*(3), 288–297.

Richardson, D., S. Wing, and W. Hoffmann (2001), Cancer risk from low-level ionizing radiation: the role of age at exposure., *Occupational medicine (Philadelphia, Pa.)*, *16*(2), 191.

Richardson, D., S. Wing, K. Steenland, and W. McKelvey (2004), Time-related aspects of the healthy worker survivor effect, *Ann Epidemiol*, *14*(9), 633–9, doi:10.1016/j.annepidem.2003.09.019.

Richardson, D. B. (2009a), Multistage modeling of leukemia in benzene workers: a simple approach to fitting the 2-stage clonal expansion model, *Am J Epidemiol*, *169*(1), 78–85, doi:10.1093/aje/kwn284.

Richardson, D. B. (2009b), Latency models for analyses of protracted exposures, *Epidemiology*, *20*(3), 395–9, doi:10.1097/EDE.0b013e318194646d.

Richardson, D. B., and J. P. Ashmore (2005), Investigating time patterns of variation in radiation cancer associations, *J Occup Environ Med*, *62*(8), 551–8, doi:10.1136/oem.2004.017368.

Richardson, D. B., S. R. Cole, H. Chu, and B. Langholz (2011), Lagging exposure information in cumulative exposure-response analyses, *Am J Epidemiol*, doi:10.1093/aje/kwr260.

Richardson, D. B., S. R. Cole, and B. Langholz (2012), Regression models for the effects of exposure rate and cumulative exposure, *Epidemiology*, *23*(6), 892–899, doi:10.1097/EDE.0b013e31826c3149.

Ries, L., M. Eisner, C. Kosary, B. Hankey, B. Miller, L. Clegg, and B. Edwards (2005), Cancer statistics review, 1975–2002, *Bethesda, MD: National Cancer Institute*, *21*.

Ringholz, R. (2002), *Uranium frenzy: Saga of the nuclear West*, Utah State Univ Pr.

Robins, J. (1999), Association, causation, and marginal structural models, *Synthese*, *121*(1), 151–179.

Robins, J. M. (1986), 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*, *7*(9), 1393–1512.

Robins, J. M. (1987a), Addendum to a new approach to causal inference in mortality studies with a sustained exposure period application to control of the healthy worker survivor effect, *Comput Math Appl*, *14*(912), 923–945.

Robins, J. M. (1987b), A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods, *J Chron Dis*, *40*(S2), 139S–161S.

Robins, J. M. (1989), *The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: Sechrest L, Freeman H, Mulley A, eds. Health service research methodology: a focus on AIDS*, 113-159 pp., National Center for Health Services Research, US Public Health Service.

Robins, J. M. (1992), Estimation of the time-dependent accelerated failure time model in the presence of confounding factors, *Biometrika*, *79*(2), 321.

Robins, J. M. (1997), Marginal structural models, Alexandria, VA, 1-10 pp.

Robins, J. M. (1998), Structural nested failure time models, *Survival Analysis: Encyclopedia of Biostatistics*.

Robins, J. M., and A. Tsiatis (1992), Semiparametric estimation of an accelerated failure time model with time-dependent covariates, *Biometrika*, *79*(2), 311–319.

Robins, J. M., and L. Wasserman (1997), Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs, *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, pp. 409–420.

Robins, J. M., D. Blevins, G. Ritter, and M. Wulfsohn (1992), G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of aids patients, *Epidemiology*, *3*(4), 319–36.

Robins, J. M., M. A. Hernán, and B. A. Brumback (2000), Marginal structural models and causal inference in epidemiology, *Epidemiology*, *11*(5), 550–560.

Robins, J. M., M. A. Hernán, and U. Siebert (2004), Effects of multiple interventions, *Comparative Quantification of Health Risks: The Global and Regional Burden of Disease Attributable to Major Risk Factors, Genève, Organisation mondiale de la Santé*.

Roscoe, R. (1997), An update of mortality from all causes among white uranium miners from the colorado plateau study group, *Am J Ind Med*, *31*(2), 211–222.

Roscoe, R. J., K. Steenland, W. E. Halperin, J. J. Beaumont, and R. J. Waxweiler (1989), Lung cancer mortality among nonsmoking uranium miners exposed to radon daughters, *JAMA*, *262*(5), 629–33.

Roscoe, R. J., J. A. Deddens, A. Salvan, and T. M. Schnorr (1995), Mortality among navajo uranium miners, *Am J Public Health*, *85*(4), 535–40, doi:10.2105/AJPH.85.4.535.

Rosenbaum, P. (1984), The consquences of adjustment for a concomitant variable that has been affected by the treatment, *J R Stat Soc Ser A-G*, *147*, 656–666.

Rosenbaum, P. R. (2007), Interference between units in randomized experiments, *J Am Stat Assoc*, *102*(477).

Rothman, K. J. (1981), Induction and latent periods, *Am J Epidemiol*, *114*(2), 253–9.

Rothman, K. J., S. Greenland, and T. Lash (2008), *Modern epidemiology*, 3nd ed., Lippincott, Williams, & Wilkins, Philadelphia, PA.

Sathiakumar, N., E. Delzell, and O. Abdalla (1998), Using the national death index to obtain underlying cause of death codes, *J Occup Environ Med*, *40*(9), 808–13.

Schiager, K. J. (1989), Bias in united states u miners' exposure data, *Health Phys*, *57*(1), 169–70.

Schnelzer, M., G. P. Hammer, M. Kreuzer, A. Tschense, and B. Grosche (2010), Accounting for smoking in the radon-related lung cancer risk among german uranium miners: results of a nested case-control study, *Health Phys*, *98*(1), 20–8, doi:10.1097/HP.0b013e3181b8ce81.

Schubauer-Berigan, M. K., R. D. Daniels, and L. E. Pinkerton (2009), Radon exposure and mortality among white and american indian uranium miners: an update of the colorado plateau cohort, *Am J Epidemiol*, *169*(6), 718–30, doi:10.1093/aje/kwn406.

Schüttmann, W. (1993), Schneeberg lung disease and uranium mining in the saxon ore mountains (erzgebirge), *Am J Ind Med*, *23*(2), 355–68.

Sevc, J., L. Tomásek, E. Kunz, V. Placek, D. Chmelevsky, D. Barclay, and A. M. Kellerer (1993), A survey of the czechoslovak follow-up of lung cancer mortality in uranium miners, *Health Phys*, *64*(4), 355–69.

Slattery, M. L. (2004), Physical activity and colorectal cancer, *Sports Med*, *34*(4), 239–52.

Stayner, L., K. Steenland, M. Dosemeci, and I. Hertz-Picciotto (2003), Attenuation of exposure-response curves in occupational cohort studies at high exposure levels, *Scand J Work Environ Health*, *29*(4), 317–24.

Steck, D. J. (2012), The effectiveness of mitigation for reducing radon risk in single-family minnesota homes, *Health Phys*, *103*(3), 241–248, doi:10.1097/HP.0b013e318250c37a.

Steenland, K., and L. Stayner (1991), The importance of employment status in occupational cohort mortality studies, *Epidemiology*, *2*(6), 418–23.

Stram, D. O., B. Langholz, M. Huberman, and D. C. Thomas (1999), Correcting for exposure measurement error in a reanalysis of lung cancer mortality for the colorado plateau uranium miners cohort, *Health Phys*, *77*(3), 265–75.

Suissa, S. (2008), Immortal time bias in pharmacoepidemiology, *Am J Epidemiol*, *167*(4), 492–499.

Taubman, S. L., J. M. Robins, M. A. Mittleman, and M. A. Hernán (2009), Intervening on risk factors for coronary heart disease: an application of the parametric g-formula, *Int J Epidemiol*, *38*(6), 1599–1611, doi:10.1093/ije/dyp192.

Thiébaut, A., and J. Bénichou (2004), Choice of time-scale in cox's model analysis of epidemiologic cohort data: a simulation study, *Stat Med*, *23*(24), 3803–3820.

Thomas, D. C. (2014), Invited commentary: is it time to retire the "pack-years" variable? maybe not!, *Am J Epidemiol*, *179*(3), 299–302, doi:10.1093/aje/kwt274.

Thygesen, L. C., U. A. Hvidtfeldt, S. Mikkelsen, and H. Brønnum-Hansen (2011), Quantification of the healthy worker effect: a nationwide cohort study among electricians in denmark, *BMC Public Health*, *11*, 571, doi:10.1186/1471-2458-11-571.

Tomásek, L. (2011), Interaction of radon and smoking among czech uranium miners, *Radiat Prot Dosimetry*, *145*(2-3), 238–42, doi:10.1093/rpd/ncr048.

Tomásek, L. (2012), Lung cancer mortality among czech uranium miners-60 years since exposure, *J Radiol Prot*, *32*(3), 301–314, doi:10.1088/0952-4746/32/3/301.

Tomásek, L., A. Rogel, M. Tirmarche, N. Mitton, and D. Laurier (2008), Lung cancer in french and czech uranium miners: Radon-associated risk at low exposure rates and modifying effects of time since exposure and age at exposure, *Radiat Res*, *169*(2), 125–37, doi:10.1667/RR0848.1.

UNSCEAR (2008), *UNSCEAR 2006 Report: Effects of Ionizing Radiation: Report to the General Assembly, with scientific annexes*, vol. 1, United Nations Publications.

USEPA (2003), Report: Epa's assessment of risks from radon in homes.

Vacquier, B., A. Rogel, K. Leuraud, S. Caer, A. Acker, and D. Laurier (2009), Radon-associated lung cancer risk among french uranium miners: modifying factors of the exposure–risk relationship, *Radiation and environmental biophysics*, *48*(1), 1–9.

van der Laan, M. J., E. C. Polley, and A. E. Hubbard (2007), Super learner, *Tech. Rep. 222*, Division of Biostatistics, University of California, Berkeley.

VanderWeele, T. J. (2009), Concerning the consistency assumption in causal inference, *Epidemiology*, *20*(6), 880–883, doi:10.1097/EDE.0b013e3181bd5638.

VanderWeele, T. J., and S. Vansteelandt (2011), A weighting approach to causal effects and additive interaction in case-control studies: marginal structural linear odds models, *Am J Epidemiol*, *174*(10), 1197–203, doi:10.1093/aje/kwr334.

Vlaanderen, J., et al. (2014), Effect modification of the association of cumulative exposure and cancer risk by intensity of exposure and time since exposure cessation: a flexible method applied to cigarette smoking and lung cancer in the synergy study, *Am J Epidemiol*, *179*(3), 290–8, doi:10.1093/aje/kwt273.

Wagoner, J., V. Archer, B. Carroll, D. Holaday, and P. Lawrence (1964), Cancer mortality patterns among us uranium miners and millers, 1950 through 1962, *J Natl Cancer I*, *32*.

Walsh, L., F. Dufey, A. Tschense, M. Schnelzer, B. Grosche, and M. Kreuzer (2010a), Radon and the risk of cancer mortality–internal poisson models for the german uranium miners cohort, *Health Phys*, *99*(3), 292–300, doi:10.1097/HP.0b013e3181cd669d.

Walsh, L., A. Tschense, M. Schnelzer, F. Dufey, B. Grosche, and M. Kreuzer (2010b), The influence of radon exposures on lung cancer mortality in german uranium miners, 1946-2003, *Radiat Res*, *173*(1), 79–90, doi:10.1667/RR1803.1.

Weinberg, C. R. (1993), Toward a clearer definition of confounding, *Am J Epidemiol*, *137*(1), 1–8.

Westreich, D., and S. R. Cole (2010), Invited commentary: positivity in practice, *Am J Epidemiol*, *171*(6), 674–677, doi:10.1093/aje/kwp436.

Westreich, D., S. R. Cole, J. G. Young, F. Palella, P. C. Tien, L. Kingsley, S. J. Gange, and M. A. Hernán (2012), The parametric g-formula to estimate the effect of highly active antiretroviral therapy on incident aids or death, *Stat Med*, *31*(18), 2000–2009.

Whittemore, A. S., and A. McMillan (1983), Lung cancer mortality among u.s. uranium miners: a reappraisal, *J Natl Cancer Inst*, *71*(3), 489–99.

Witteman, J. C., R. B. D'Agostino, T. Stijnen, W. B. Kannel, J. C. Cobb, M. A. de Ridder, A. Hofman, and J. M. Robins (1998), G-estimation of causal effects: isolated systolic hypertension and cardiovascular death in the framingham heart study, *Am J Epidemiol*, *148*(4), 390–401.

World Health Organization (2009), *WHO handbook on indoor radon: a public health perspective*, Geneva: World Health Organization.

Yarmoshenko, I. V., A. V. Vasilyev, A. D. Onishchenko, S. M. Kiselev, and M. V. Zhukovsky (2014), Indoor radon problem in energy efficient multi-storey buildings, *Radiat Prot Dosimetry*, doi:10.1093/rpd/ncu110.

Young, J., M. A. Hernán, S. Picciotto, and J. M. Robins (2010), Relation between three classes of structural models for the effect of a time-varying exposure on survival, *Lifetime data anal*, *16*(1), 71–84.

Zhukovsky, M. V., and A. V. Vasilyev (2014), Mechanisms and sources of radon entry in buildings constructed with modern technologies., *Radiat Prot Dosimetry*, doi:10.1093/rpd/ncu111.