

ORIGINAL ARTICLE

Assessment of the healthy worker survivor effect in cancer studies of the United Autoworkers-General Motors cohort

Erika Garcia,¹ Sally Picciotto,¹ Sadie Costello,¹ Patrick T Bradshaw,² Ellen A Eisen¹

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/oemed-2016-104038>)

¹Environmental Health Sciences Division, School of Public Health, University of California, Berkeley, California, USA

²Epidemiology Division, School of Public Health, University of California, Berkeley, California, USA

Correspondence to

Ellen A Eisen, Environmental Health Sciences Division, School of Public Health, University of California, 50 University Hall #7360, Berkeley, CA 94720-7360, USA; eeisen@berkeley.edu

Received 23 August 2016
Revised 26 November 2016
Accepted 19 December 2016
Published Online First
9 January 2017

ABSTRACT

Objective The healthy worker survivor effect (HWSE) can affect the validity of occupational studies when data are analysed incorrectly. HWSE depends on three underlying conditions: (1) leaving work predicts future exposure, (2) leaving work is associated with disease outcome and (3) prior exposure increases probability of leaving work. If all these conditions are satisfied, then employment status is a time-varying confounder affected by prior exposure, and standard regression will produce bias. We assessed these conditions for cancer outcomes in a cohort of autoworkers exposed to metalworking fluids (MWF).

Methods The cohort includes 31 485 workers followed for cancer incidence from 1985 to 1994. As occupational exposures to straight, soluble and synthetic MWFs are necessarily zero after leaving work, condition (1) is satisfied. Cox models for cancer incidence and for employment termination were used to assess conditions (2) and (3), respectively. Employment termination by select ages was examined to better gauge the presence of condition (2).

Results The HR for leaving work as a predictor of all cancers combined and prostate cancer was null, but elevated for lung and colorectal cancers among men. Condition (2) was more clearly satisfied for all cancer outcomes when leaving work occurred by age 50. Higher exposures to all three MWF types were associated with increased rates of leaving work (condition (3)), with the exception of straight MWF among women.

Conclusions We found evidence for the structural conditions underlying HWSE in a cohort of autoworkers. G-methods should be applied to reduce HWSE bias in studies of all cancers presently examined.

INTRODUCTION

The healthy worker survivor effect (HWSE) can affect the validity of results from occupational health studies.^{1–6} It has been extensively studied^{1–6} and we describe it only briefly here. The HWSE arises due to less healthy workers preferentially leaving the workforce, resulting in workers who remain at work (ie, survive) being healthier than those who have left. A directed acyclic graph (DAG) is a mathematical tool widely used in epidemiology to address problems of causal inference.⁷ Introduced into epidemiology more than a decade ago, these graphs explicitly present assumptions about the temporal relations between the exposures, covariates and the outcome, allowing us

What this paper adds

- The healthy worker survivor effect (HWSE) is a well-known concern for occupational health studies and can affect the validity of results if inappropriate methods are used to analyse the data.
- The presence of the HWSE is dependent on the study cohort and the outcome of interest being examined.
- We assessed the potential for bias due to the HWSE in a longitudinal study of cancer incidence and metalworking fluid exposure and found that advanced methods are needed to address the HWSE and avoid bias in exposure–response analyses for all cancer outcomes examined.
- Examination of the three conditions for the HWSE should precede exposure–response analyses in other occupational cohort studies to assess the evidence that it is present.

to identify variables that need to be controlled in order to estimate results unbiased by confounding.⁸ In [figure 1](#) we show a DAG depicting the structure that gives rise to the HWSE. Unobserved underlying health status (*U* in [figure 1](#)) may influence leaving work (employment status at time *j*) and risk of the disease of interest. Leaving work may be affected by prior exposure (exposure at time *j*-1) and also directly affects subsequent exposure (exposure at time *j*) as occupational exposure will be zero after leaving work. In studies of cumulative exposure, employees who leave work accrue less exposure than their counterparts who remain on the job, causing bias induced by the HWSE.³

As initially described by Robins, the HWSE poses a challenge to conventional methods of analysis because employment termination status is a time-varying confounder affected by prior exposure.⁶ Employment status is associated with the disease and predicts future exposure (making it a confounder of future exposure and disease outcome) and it is also predicted by past exposure (making it an intermediate between past exposure and disease outcome). This poses a dilemma whereby we need to control for employment status as a confounder and need to *not* control for it since it is on the causal pathway between exposure



To cite: Garcia E, Picciotto S, Costello S, *et al.* *Occup Environ Med* 2017;**74**:294–300.

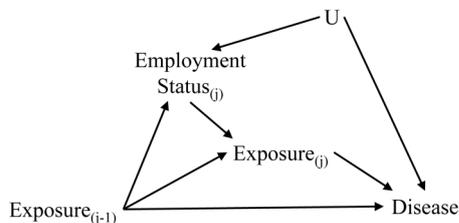


Figure 1 Directed acyclic graph depicting the associations underlying healthy worker survivor effect.

and disease outcome. Furthermore, adjusting for employment status would induce collider-stratification bias by opening a confounding pathway as past exposure and underlying health status are parents of employment status.⁹

The healthy worker survivor effect is sometimes referred to as healthy worker survivor bias.² If we define the effect as the data structure described above (depicted in figure 1), however, then we can distinguish it from the bias. Bias from this source occurs only if inappropriate analytical methods are applied. When standard methods (eg, covariate adjustment in Poisson or Cox proportional hazards regression) are applied where the HWSE is present, results will likely be biased.^{2–6} More advanced g-methods, including the g-null test,¹⁰ g-estimation of structural nested models,¹¹ the parametric g-formula^{6, 10} and inverse-probability-weighted marginal structural models,¹² have been developed to control for this type of bias.

To guide our analytical approach in a cohort study of cancer incidence and metalworking fluids (MWF) exposure, we first assessed the potential for bias due to the HWSE. We regard evidence that employment status is a time-varying confounder affected by prior exposure as an indicator of the potential bias. Thus, the presence of the HWSE depends on satisfying three underlying conditions:

1. leaving work predicts future exposure;
2. leaving work is associated with disease outcome;
3. prior exposure increases probability of leaving work.

Since occupational exposure is necessarily zero after leaving work, (1) is inherently satisfied. In this study, we assessed the presence of conditions (2) and (3), which would indicate that HWSE threatens the validity of traditional regression results in epidemiological studies of Michigan autoworkers exposed to MWF. This assessment of the presence of the HWSE will guide the choice of methodological approach in future cancer incidence studies.

METHODS

Study population

The United Autoworkers-General Motors (UAW-GM) study is an occupational cohort study designed in the mid-1980s to examine cancer mortality and its relation to MWF exposure. The original cohort includes over 46 000 hourly workers employed for at least 3 years in one of three automobile manufacturing plants in Michigan, USA, and hired before 1 January 1985.¹³ The study population was further restricted to subjects alive and aged 75 years or younger on 1 January 1985, the year follow-up of cancer incidence began via the Michigan Cancer Registry. As almost all subjects have left work by the age of 75, the age restriction was added as an eligibility criterion and for administrative censoring to ensure adequate data coverage when examining employment status. Subjects missing more than half of their work history information were excluded from the analyses (3.1%). The final study population comprised 31 485

autoworkers. Information on date of birth, year of hire, race, sex and plant were ascertained from workers' employment records.

Exposure assessment

MWFs are coolants and lubricants used in industrial machining and grinding operations and are categorised into three classes based on composition: straight (oil-based), soluble (oils emulsified in water) and synthetic (water-soluble chemical lubricants without oil). Quantitative levels of exposure to each MWF class were assigned to subjects based on detailed employment records and a time-varying job-exposure matrix (JEM).^{14, 15} The JEM was created by estimating MWF concentrations as an 8-hour time-weighted average (mg/m^3) based on 475 full-shift personal air samples collected in the mid-1980s in major exposure groups in the three manufacturing plants. A set of multipliers (scale factors) was developed to adjust MWF concentration for temporal trends based on 394 historical air sampling measurements, review of historical records and interviews with plant personnel. Scale factors were updated in 1995 after revisiting the three plants. The JEM was combined with employment records to estimate time-weighted annual exposure to each of straight, soluble and synthetic MWFs (mg/m^3). For each worker, cumulative exposures to straight, soluble and synthetic MWFs (mg/m^3 -years) were estimated by summing over annual exposures during employment.

Case ascertainment

The UAW-GM cancer incidence cohort was linked with the Michigan Cancer Registry to obtain incident cancer cases diagnosed between 1 January 1985 and 31 December 1994. Michigan cancer data are collected by the Michigan Department of Community Health as part of the Michigan Cancer Surveillance Program,¹⁶ which participates in the National Program of Cancer Registries of the Centers for Disease Control and Prevention.¹⁷ The outcomes of interest for this study are first diagnosis of any (International Classification of Disease for Oncology Third Edition codes C00.0-C80.9), prostate (C61.0-C61.9), lung (C34.0-C34.9) and colorectal (C18.0-C18.9, C19.9, C20.9, C21.1-C21.2, C21.8) cancers. The site-specific cancers were selected based on having at least 100 identified incident cases during the follow-up period. Vital status was obtained through linkage with the National Death Index (National Center for Health Statistics, Hyattsville, Maryland, USA).

Statistical methods

We assessed the presence of two conditions necessary for the HWSE by estimating HRs using Cox proportional hazards models to predict cancer incidence and employment termination, following an approach similar to Naimi *et al.*¹⁸ For all models, we used age as the time metric. All models were additionally adjusted for race, plant and calendar year (B-spline with 3 df and equally spaced knots). Separate models were fit for male and female workers due to temporal differences in hiring and exposure patterns.

For condition (2), we modelled cancer outcome as a function of employment termination (=1 if subject left work), which was treated as a time-varying indicator variable. Follow-up began on 1 January 1985 and ended on either the date of first cancer diagnosis (any or site specific), date of death, 75th birthday or 31 December 1994 (the last date for which we have employment data), whichever occurred first. Year of cohort entry (B-spline with 3 df and equally spaced knots) and duration of

employment were additionally adjusted for in the models. In order to gauge whether condition (2) is present, we must block the backdoor path⁸ leading from termination status through prior exposure to cancer incidence in figure 1. To accomplish this, we further adjusted the model for previous exposure using cumulative exposure accrued up through the prior year. Owing to cancer latency time, we do not believe exposure in the current year would affect cancer diagnosis in that same year. As such, we did not include current exposure in the cancer incidence models. This differs from the methods used by Naimi *et al*¹⁸ who were examining lung cancer mortality and did adjust for subsequent exposure.

We also investigated whether an age-specific measure of leaving work better distinguishes employment termination for health-related reasons (at younger ages) from normal retirement (at older ages). Our hypothesis was that leaving work at a younger age would be a stronger predictor of future cancer risk than leaving at any age. To explore this hypothesis, we examined condition (2) based on leaving work by age 50, 55 and 60. Instead of a single time-varying indicator of employment termination, a set of two dummy variables was constructed to capture three levels of employment status: actively employed workers, subjects who left work by the age cut point and subjects who left work after. This construction allowed the main model with no age cut point and these age-specific models to have a common reference group (ie, actively employed workers).

For assessing condition (3), cumulative MWF exposure was accrued up through the prior year and parameterised as a categorical variable with cut points defined based on the exposure distribution among subjects who had left work. Separate terms for each MWF class were included simultaneously in all the models. For these models we additionally adjusted for year of hire. Follow-up for leaving work began when subjects entered the cohort (3 years after date of hire) and ended on either the date of employment termination or 31 December 1994, whichever occurred first. We also examined cumulative exposure lagged by 15 years in our evaluation of condition (3), because we plan to lag exposure in the ultimate exposure–response models for cancer incidence. This approach is also consistent with an earlier proposal for reducing the HWSE by lagging exposures in order to ignore the more recent exposures most likely to affect leaving work.⁵

SAS software V9.4 (SAS Institute, Cary, North Carolina, USA) was used for all analyses. All research protocols were approved by the Office for the Protection of Human Subjects at the University of California at Berkeley.

RESULTS

Characteristics of the 31 485 autoworkers included in the cancer incidence cohort are presented in table 1. These characteristics are shown according to the two different follow-up periods: for cancer incidence of any site in the assessment of condition (2) and for leaving work in the assessment of condition (3). The cohort is predominately white men, but includes 19% (N=5972) African-American and 13% (N=4228) female workers. The person-years for both follow-up periods have similar distributions for most covariates. Workers were more likely to be exposed to soluble MWF than either of the two other fluid types. The exposure level among exposed workers was highest for soluble, measured as annual and cumulative exposure. Male workers were hired earlier than female workers (median (IQR): 1966 (1953–1974) vs 1975 (1967–1977)), at younger ages (23 (19–28) vs 27 (22–34)), and were

Table 1 Demographic and exposure characteristics of 31 485 autoworkers in the United Autoworkers-General Motors incidence cohort who were alive in 1985, during follow-up for cancer incidence and follow-up for leaving work (median (IQR), unless otherwise noted)*

Characteristic	During cancer follow-up† (starting in 1985)	During employment follow-up (starting at date of hire‡)
N subjects	31 485	31 485
Person-years	286 023	504 836
Age	46 (38–58)	38 (31–46)
Sex, n (%)		
Male	247 220 (86.4)	447 576 (88.7)
Female	38 803 (13.6)	57 260 (11.3)
Race, n (%)		
White	231 186 (80.8)	399 730 (79.2)
African-American	54 837 (19.2)	105 106 (20.8)
Year of hire	1968 (1959–1976)	1966 (1953–1973)
Calendar year	1989 (1987–1992)	1980 (1970–1986)
At work, n (%)	160 242 (56.0)	504 836 (100)
Employment terminations, n	–	20 586
Incident cancer cases, n	1739	–
MWF exposure§		
Any straight, n (%)	24 367 (8.5)	125 287 (24.8)
Straight, mg/m ³	0.09 (0.04–0.25)	0.10 (0.04–0.39)
Cumulative straight, mg/m ³ -years	0.64 (0.21–2.33)	0.72 (0.22–2.85)
Any soluble, n (%)	78 148 (27.3)	214 299 (42.4)
Soluble, mg/m ³	0.22 (0.13–0.34)	0.37 (0.18–0.87)
Cumulative soluble, mg/m ³ -years	3.77 (1.54–9.54)	4.31 (1.54–11.51)
Any synthetic, n (%)	24 380 (8.5)	83 810 (16.6)
Synthetic, mg/m ³	0.05 (0.02–0.17)	0.04 (0.02–0.17)
Cumulative synthetic, mg/m ³ -years	0.38 (0.15–1.33)	0.38 (0.13–1.39)

* All n statistics presented as number of person-years, except for employment terminations and incident cancer cases which are presented as number of subjects.

† Follow-up for first diagnosis of any cancer site.

‡ Start of follow-up is date of hire+3 years.

§ Annual and cumulative exposure levels reported among those exposed.

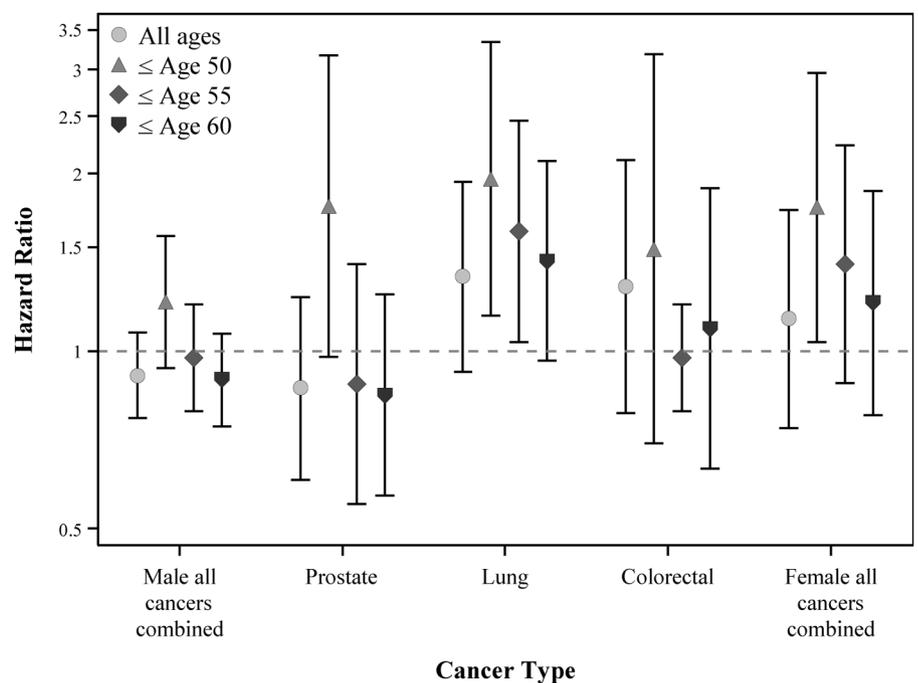
employed at the company for more years (19 (11–28) vs 17 (12–20)). Annual and cumulative exposures were higher for male than female workers for all fluid types, except for annual synthetic exposure, which was similar between the two groups. Owing to the differences in work and exposure patterns among male and female workers, all analyses were stratified by sex.

As seen in table 2, among the 1739 incident cases of a first cancer diagnosed at any site, female cancer cases were younger at diagnosis and more likely to be actively employed than male cancer cases. For women, no site-specific cancer had more than 100 cases and so none were assessed. For men, we focused on the three most common cancers: prostate, lung and colorectal. Prostate cancer cases were diagnosed at older ages than the other site-specific cancers, and survival was worst for lung cancer with regard to case fatality and survival time. Colorectal cancer cases were more likely to be diagnosed while actively employed, whereas prostate cancer cases were least likely to be actively employed at the time of diagnosis.

The results of the assessment for condition (2) were not entirely consistent across the five cancer outcomes examined (figure 2). Based on the main model adjusted for covariates and past exposure, we found the HR associated with having left

Table 2 Demographic characteristics of incident cancer cases diagnosed 1985–1994 among autoworkers in the United Autoworkers-General Motors incidence cohort who were alive in 1985 (median (IQR), unless otherwise noted)

	Male workers				Female workers
	All cancers combined	Prostate cancer	Lung cancer	Colorectal cancer	All cancers combined
N	1517	384	340	182	222
Year of birth	1926 (1920–1933)	1923 (1919–1929)	1925 (1920–1932)	1925 (1919–1931)	1928 (1920–1942)
Year entered cohort	1956 (1952–1969)	1956 (1951–1966)	1956 (1952–1967)	1956 (1951–1969)	1970 (1955–1979)
Race, n (%)					
White	1178 (77.7)	266 (69.3)	265 (77.9)	144 (79.1)	179 (80.6)
African-American	339 (22.4)	118 (30.7)	75 (22.1)	38 (20.9)	43 (19.4)
Age at diagnosis	65.1 (57.3–70.4)	68.2 (63.1–72)	65.5 (57.5–71.5)	65.5 (59.5–69.8)	62.4 (50.1–70)
Deceased, n (%)	1276 (84.1)	286 (74.5)	336 (98.8)	155 (85.2)	161 (72.5)
Survival among deceased, years	1.4 (0.1–6.9)	6.5 (2.6–11.9)	0.2 (0–1.1)	2.7 (0.7–10.2)	1.5 (0.3–8.9)
Termination status, n (%)					
At work	394 (26)	67 (17.5)	65 (19.1)	39 (21.4)	81 (36.5)
Left work	1123 (74)	317 (82.6)	275 (80.9)	143 (78.6)	141 (63.5)
Age left work	54.9 (36.4–61.8)	57.4 (43.7–62.3)	52.4 (36.5–60.7)	57.1 (35.7–62)	49.4 (37.5–58.2)

Figure 2 Condition 2: leaving working and cancer incidence. Site-specific cancers examined only among male workers. HRs with 95% CIs. All modelled as Cox regression with age as time scale and adjusted for race, plant, year of cohort entry, calendar year, duration of employment and prior exposure.

work was null for all cancers combined and for prostate cancer, slightly elevated for colorectal cancers and elevated with borderline significance for lung cancer. However, when we considered the age at having left work, the HRs increased for all outcomes.

Based on magnitude and statistical significance, at least one category of MWF exposure was associated with leaving work for all fluid types among men (table 3). Compared with lowest exposed men, those in the second quartile of cumulative exposure to oil-based straight and soluble MWFs were at a higher risk for leaving work. In contrast, only those in the highest exposure category of synthetic MWFs were at a higher risk for leaving work. Among female workers, MWF exposure was associated with leaving work only for the water-based soluble and synthetic fluid types. Examining condition (3) using exposure lagged by 15 years did not change the overall conclusion that MWF exposure is associated with leaving work in male and female workers (data not shown).

DISCUSSION

The aim of this study was to assess the potential for bias due to the HWSE in a longitudinal study of cancer incidence and MWF exposure in order to guide our future analytical approach. The first condition required for the HWSE is present by definition: (1) leaving work predicts future MWF exposure, because in this, as in all occupational cohort studies, workers cannot be occupationally exposed after employment termination. One caveat is that after terminating work at one of the three study plants, subjects may have been hired at another automobile manufacturing plant where MWF were also in use. Although leaving work is not necessarily an indicator of zero future exposure, we have assumed that it is. This is a limitation of the study, as described below.

Evidence of a higher risk of all cancer outcomes was observed for those who left work by age 50 (condition (2)). Across all outcomes leaving work by age 50 was a stronger predictor of

Table 3 Condition 3: cumulative exposure to MWF and leaving work among male and female autoworkers*

Male workers (N=27 257)				Female workers (N=4228)			
MWF exposure† (mg/m ³ -years)	N left work	Person-years	Adjusted HR (95% CI)	MWF exposure† (mg/m ³ -years)	N left work	Person-years	Adjusted HR (95% CI)
Straight				Straight			
0	7979	206 053.5	Ref	0	1126	30 432.5	Ref
0–0.38482	3374	85 952.0	1.13 (1.08 to 1.18)	0–0.180	452	9333.3	1.02 (0.90 to 1.15)
0.38482–2.058	3375	81 198.7	1.05 (1.00 to 1.10)	0.180–0.963	453	10 108.7	0.95 (0.84 to 1.09)
>2.058	3374	74 371.5	1.03 (0.98 to 1.07)	>0.963	453	7385.3	1.01 (0.89 to 1.15)
Soluble				Soluble			
0–1.2742	4625	140 658.5	Ref	0–0.11	621	18 314.0	Ref
1.2742–4.658	4526	115 882.0	1.16 (1.11 to 1.21)	0.11–1.522	621	16 836.4	1.06 (0.94 to 1.19)
4.658–14.573	4526	111 218.9	1.01 (0.97 to 1.06)	1.522–4.07	621	12 636.9	1.13 (0.99 to 1.28)
>14.573	4525	79 816.3	1.01 (0.96 to 1.06)	>4.07	621	9472.4	0.98 (0.86 to 1.13)
Synthetic				Synthetic			
0	12 652	314 853.2	Ref	0	1604	39 879.6	Ref
0–0.2555	1816	54 235.8	0.91 (0.86 to 0.97)	0–0.167	293	6945.7	1.15 (0.99 to 1.33)
0.2555–1.464	1817	46 134.1	0.95 (0.89 to 1.01)	0.167–0.81	294	6076.7	1.12 (0.96 to 1.31)
>1.464	1817	32 352.7	1.08 (1.02 to 1.15)	>0.81	293	4357.8	1.11 (0.96 to 1.29)

* Modelled stratified by sex using Cox models with age as time scale and adjusted for race, plant, calendar year and year of hire. All MWF types were included in the same model.

† Upper bound included in range.

MWF, metalworking fluids.

cancer than leaving work without the age restriction. This pattern is consistent with our hypothesis that workers who terminate employment prematurely are more likely to have a higher risk of the disease outcome. Evidence for condition (2) was stronger for lung and prostate cancer among men, and all cancers combined among women, than for colorectal and all cancers combined among men. Although we cannot state for certain whether this association is due to a causal link between leaving work and these cancers or by an unmeasured confounder, we believe it is most likely the latter. An unmeasured factor, such as health status, could lead to increased risk of employment termination (eg, preferential work termination by less healthy workers) and cancer. We did observe younger age at diagnosis among those who left work by age 50 compared with those who left work after age of 50 for prostate, lung, colorectal and all cancers combined among men (range 3.8–4.8 years earlier). For all cancers combined among women, this difference was only 1.4 years.

We do not have any information on why subjects in this cohort left work. Ideally, we would separate employee-initiated termination (more likely to be due to workers' underlying health status) from lay-offs or other company-initiated terminations. In the absence of any way to distinguish these, we assume all employment termination was voluntary and serves as a proxy for underlying health status. Together with condition (1), these results suggest that employment status is a time-varying confounder and researchers should control for it in their analyses.

We found evidence for the third condition required for the HWSE among all workers: (3) prior MWF exposure increases probability of leaving work. In our analyses, exposure to each of the three types of MWF (straight, soluble and synthetic) was associated with leaving work among men, and exposure to water-based soluble and synthetic MWF were each associated with leaving work among women. Lagging exposure did not alter these associations and, therefore, would not be an effective strategy to avoid bias due to the HWSE. The presence of conditions (1) and (3) imply that employment status should not be

adjusted for in our cancer risk models because it is an intermediate between exposure and disease outcome.¹⁸

The evidence for all three conditions indicates that employment status is a time-varying confounder (conditions (1) and (2)) affected by prior exposure (condition (3)). We need to control for employment status as a confounder and to not control for it since it is an intermediate. The presence of the HWSE does not in itself dictate that results will be biased; it is the analytical method applied in the presence of the HWSE that determines whether results will be biased. The observed results for the three conditions suggest that g-methods should be used to address the HWSE and avoid bias in exposure–response analyses for all cancers outcomes examined.¹⁸ G-methods were originally developed to address the HWSE.¹⁰ This is achieved by unpacking cumulative exposure into a series of time-varying exposures and building separate models for outcome, exposure and sometimes other covariates. A large literature on this class of analytical methods explains how these additional models are used to adjust properly for time-varying confounding (by leaving work) and estimate marginal differences in survival within a counterfactual framework.^{1 2 6 10 12 19}

These results are generally consistent with findings from a 2012 study of this same cohort comparing standard Cox models with g-estimation models for straight MWF and several cancer mortality outcomes.¹¹ Chevrier *et al* reported statistically significant results from g-estimation, but null results using standard methods for lung and prostate cancer mortality, suggesting the presence of the HWSE. For rectal cancer mortality, standard methods produced null results, while elevated, though non-significant, risks were found using g-estimation. There was no marked difference between standard and g-estimation results for colon cancer mortality. Costello *et al*²⁰ reported less evidence for the HWSE among women compared with men in an ischaemic heart disease analysis of this same cohort. Differences in the evidence for the HWSE between men and women in the present study are difficult to distinguish since small numbers prevented the examination of site-specific cancers among female workers.

We presented one DAG for HWSE. Slight variations of this DAG have also been described.¹⁸ Rather than a direct arrow from prior exposure to employment status, there could be, instead or in addition, an unmeasured factor leading to both of these nodes, thereby inducing an association. Similarly, where we have a U leading to employment status and disease there could be, instead or in addition, a direct arrow leading from employment status to the outcome. Each of these DAGs contains a time-varying confounder affected by prior exposure and its presence can be assessed using the methods presented here. See Naimi *et al*¹⁸ online supplementary appendix A1 for details.

Underlying health status may affect exposure in several ways: less healthy employees may take more intermittent time off work, transfer to lower exposed jobs, use personal protective equipment or terminate employment earlier. In our assessment of the conditions for the HWSE, we only explored the association through the latter pathway. An analysis that takes into account these additional potential pathways between exposure and underlying health status may provide a more accurate assessment of the potential for bias due to HWSE. However, we expect terminating employment to be the strongest source of healthy worker survivor bias, assuming those who left work were not subsequently exposed elsewhere. This is because the other actions would lead to lower but not necessarily zero exposure, and because a worker who terminates employment is likely to be in worse health than one who manages to transfer to a lower exposed job at the plant. A last notable limitation for our study is the small amount of covariate information. Owing to this lack of data, we were unable to account for other potential confounders of the relationships examined here, such as smoking status or other lifestyle factors.

It has been claimed that the HWSE is unlikely to have an effect on cancer outcomes, particularly cancers with late-stage symptoms and poor survival, such as lung cancer.²¹ In the current study of MWF exposed workers, however, we found evidence of the HWSE in relation to incidence of lung cancer. By comparing results from g-methods and standard methods, the presence of the HWSE has also been demonstrated in other occupational studies of lung cancer in relation to arsenic,^{10 22} diesel,²³ radon^{24 25} and asbestos,^{26 27} as well as MWFs.^{11 28} Coupled with the theoretical underpinnings of the HWSE, these studies provide evidence that the HWSE can affect the study of cancer outcomes in occupational settings and should be taken into account when considering different methodological approaches.

There are several g-methods available for an exposure–responses analysis of MWF and cancer incidence, though not all would be appropriate for this study. Follow-up continues after subjects leave work when exposure is zero. This violates the positivity assumption and precludes the use of inverse probability of exposure weighting.¹⁹ Accelerated failure time models used with g-estimation assume the outcome is inevitable and assess whether exposure accelerates time to the event. When the outcome is rare, as is cancer, this assumption may not hold.²⁹ The parametric g-formula^{6 10} is more appropriate for rare outcomes and can incorporate quantitative exposure and dynamic treatment regimes, though it requires many parametric assumptions. When applying any of these methods in this study we will need to adjust for leaving work by a particular age (eg, 50 years), rather than leaving work regardless of age. Based on our results, this more refined variable definition for leaving work better captures the essence of the HWSE.

CONCLUSION

We found evidence for all three conditions necessary for the HWSE to exist in a cohort study of Michigan autoworkers exposed to MWFs. Evidence was strongest for leaving work by age 50, consistent with the HWSE. Lagging exposure by 15 years, in keeping with the latency period for the cancer outcomes, did not alter these results. This suggests that standard methods may underestimate the exposure–response for these outcomes and therefore a g-method should be applied to control for employment status as a time-varying confounder affected by prior exposure.^{6 10} Examination of the three conditions for the HWSE should precede exposure–response analyses in other occupational cohort studies to assess the evidence that it is present.

Contributors EG and EAE had the initial idea for this study. EG conducted data analyses, interpreted results and wrote and revised the manuscript. EAE contributed to the design of the analysis, interpretation of results and reviewed and revised the manuscript. SP and SC advised on analytical approach and reviewed and revised the manuscript. PTB reviewed and revised the manuscript. All authors read and approved the final manuscript.

Funding This research was supported by a National Institute of Occupational Safety and Health (NIOSH) Targeted Research Training grant T420H008429 and NIOSH grant R01OH010028.

Competing interests None declared.

Ethics approval Office for the Protection of Human Subjects at the University of California at Berkeley.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- 1 Arrighi HM, Hertz-Picciotto I. The evolving concept of the healthy worker survivor effect. *Epidemiology* 1994;5:189–96.
- 2 Buckley JP, Keil AP, McGrath LJ, *et al*. Evolving methods for inference in the presence of healthy worker survivor bias. *Epidemiology* 2015;26:204–12.
- 3 Picciotto S, Hertz-Picciotto I. Commentary: healthy worker survivor bias: a still-evolving concept. *Epidemiology* 2015;26:213–15.
- 4 Fox AJ, Collier PF. Low mortality rates in industrial cohort studies due to selection for work and survival in the industry. *Br J Prev Soc Med* 1976;30:225–30.
- 5 Gilbert ES. Some confounding factors in the study of mortality and occupational exposures. *Am J Epidemiol* 1982;116:177–88.
- 6 Robins J. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *J Chronic Dis* 1987;40(Suppl 2):1395–615.
- 7 Pearl J, Glymour M, Jewell NP. *Causal inference in statistics: a primer*. West Sussex, UK: Wiley, 2016.
- 8 Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;10:37–48.
- 9 Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology* 2003;14:300–6.
- 10 Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Model* 1986;7:1393–512.
- 11 Chevri er J, Picciotto S, Eisen EA. A comparison of standard methods with g-estimation of accelerated failure-time models to address the healthy-worker survivor effect: application in a cohort of autoworkers exposed to metalworking fluids. *Epidemiology* 2012;23:212–19.
- 12 Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;11:550–60.
- 13 Eisen EA, Tolbert PE, Monson RR, *et al*. Mortality studies of machining fluid exposure in the automobile industry I: a standardized mortality ratio analysis. *Am J Ind Med* 1992;22:809–24.
- 14 Hallock MF, Smith TJ, Woskie SR, *et al*. Estimation of historical exposures to machining fluids in the automotive industry. *Am J Ind Med* 1994;26:621–34.
- 15 Woskie SR, Smith TJ, Hallock MF, *et al*. Size-selective pulmonary dose indices for metal-working fluid aerosols in machining and grinding operations in the automobile manufacturing industry. *Am Ind Hyg Assoc J* 1994;55:20–9.
- 16 Michigan Department of Health and Human Services. Michigan Cancer Surveillance Program. 2015. http://www.michigan.gov/mdch/0,4612,7-132-2945_5221-16586-,00.html (accessed 5 May 2015).
- 17 Division of Cancer Prevention and Control National Center for Chronic Disease Prevention and Health Promotion. *National Program of Cancer Registries*. Centers for Disease Control and Prevention, 2014.

- 18 Naimi AI, Cole SR, Hudgens MG, *et al.* Assessing the component associations of the healthy worker survivor bias: occupational asbestos exposure and lung cancer mortality. *Ann Epidemiol* 2013;23:334–41.
- 19 Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health* 2006;60:578–86.
- 20 Costello S, Picciotto S, Rehkopf DH, *et al.* Social disparities in heart disease risk and survivor bias among autoworkers: an examination based on survival models and g-estimation. *Occup Environ Med* 2015;72:138–44.
- 21 Choi BC. Definition, sources, magnitude, effect modifiers, and strategies of reduction of the healthy worker effect. *J Occup Med* 1992;34:979–88.
- 22 Arrighi HM, Hertz-Picciotto I. Controlling the healthy worker survivor effect: an example of arsenic exposure and respiratory cancer. *Occup Environ Med* 1996;53:455–62.
- 23 Neophytou AM, Picciotto S, Costello S, *et al.* Occupational diesel exposure, duration of employment, and lung cancer: an application of the parametric G-formula. *Epidemiology* 2016;27:21–8.
- 24 Edwards JK, McGrath LJ, Buckley JP, *et al.* Occupational radon exposure and lung cancer mortality: estimating intervention effects using the parametric g-formula. *Epidemiology* 2014;25:829–34.
- 25 Keil AP, Richardson DB, Troester MA. Healthy worker survivor bias in the Colorado Plateau uranium miners cohort. *Am J Epidemiol* 2015;181:762–70.
- 26 Cole SR, Richardson DB, Chu H, *et al.* Analysis of occupational asbestos exposure and lung cancer mortality using the g formula. *Am J Epidemiol* 2013;177:989–96.
- 27 Naimi AI, Cole SR, Hudgens MG, *et al.* Estimating the effect of cumulative occupational asbestos exposure on time to lung cancer mortality: using structural nested failure-time models to account for healthy-worker survivor bias. *Epidemiology* 2014;25:246–54.
- 28 Picciotto S, Brown DM, Chevrier J, *et al.* Healthy worker survivor bias: implications of truncating follow-up at employment termination. *Occup Environ Med* 2013;70:736–42.
- 29 Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran ME, Berry DA, eds. *Statistical models in epidemiology, the environment and clinical trials*. New York: Springer, 2000;95–133.