ORIGINAL ARTICLE

Healthy worker survivor bias: implications of truncating follow-up at employment termination

Sally Picciotto,[1] Daniel M Brown,[1,2] Jonathan Chevrier,[1,3] Ellen A Eisen[1,3]

[1]Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, California, USA
[2]Division of Biostatistics, School of Public Health, University of California, Berkeley, California, USA
[3]Division of Epidemiology, School of Public Health, University of California, Berkeley, California, USA

Correspondence to
Dr Sally Picciotto, Division of Environmental Health Sciences, UC Berkeley School of Public Health, 789 University Hall, Berkeley, CA 94720, USA; sallypicciotto@berkeley.edu

Received 19 December 2012
Revised 2 July 2013
Accepted 5 July 2013
Published Online First 19 July 2013

ABSTRACT

Objectives The healthy worker survivor effect is a bias that occurs in occupational studies when less healthy workers are more likely to reduce their workplace exposures. When variables on the pathway from health status to exposure are measured, g-methods can avoid this bias. However, studies in which follow-up ends at employment termination have additional potential for selection bias. This paper examines the structure of the healthy worker survivor effect, compares results with and without censoring at employment termination, and addresses how to prevent bias when such censoring occurs.

Methods G-estimation of structural accelerated failure time models was applied in the United Autoworkers—General Motors cohort study to examine relationships between metalworking fluid exposure and cause-specific mortality. Subjects were followed from hire through 1994, regardless of employment status. To answer the central question, g-estimation analysis was repeated after truncating at employment termination and censoring outcomes that occurred thereafter, with adjustment for censoring by inverse probability weighting.

Results Using full follow-up time, HRs were estimated for all-cause mortality (1.09), ischaemic heart disease death (1.19), and death from any cancer (1.09), comparing 5 years of metalworking fluid exposure to no exposure. For all three outcomes, the HR estimates based on data censored at termination of employment were below 1 (respectively, 0.92, 0.97, 0.79).

Conclusions In this application, g-estimation together with weighting did not prevent selection bias due to employment termination. However, the bias might be avoided in studies with measured health-related variables on the pathway from health status to employment termination.

What this paper adds

- Previous studies have shown that when variables on the pathway from health status to exposure are measured and follow-up extends past termination of employment, g-methods can avoid the Healthy Worker Survivor Effect.
- However, in many occupational studies, follow-up ends at termination of employment, necessitating additional adjustment.
- This pedagogical paper illustrates the problem by comparing results from the same cohort under two scenarios: one based on data with complete follow-up and another with follow-up that ends at employment termination.
- Even using g-methods, additional assumptions and measurements of more variables may be needed to mitigate healthy worker survivor bias when follow-up ends at termination of employment.
- Causal diagrams shed light on the reasons for bias and clarify the data needed to avoid such bias.

In occupational epidemiology, the healthy worker effect refers to a well-documented downward bias—towards the null and sometimes beyond—that obscures evidence for the harmful effects of a workplace exposure. It occurs in studies of the health effects of cumulative exposure because the most robust workers accumulate more exposure than less healthy workers or the unemployed. The phenomenon has two components: the healthy hire effect, which only arises when workers are compared with external populations, and the healthy worker survivor effect (HWSE), which arises even in internal comparisons. The HWSE occurs when workers reduce their workplace exposures for health-related reasons, whether or not exposure affects health. This bias can occur whether or not follow-up extends beyond employment termination.[1 2]

Conventional methods for adjusting for HWSE include restricting analysis to active workers who survive at least 15 years after hire,[3] lagging the exposure[4] or adjusting for current employment status.[5] Steenland *et al*'s[6] simulation examined several scenarios in which the HWSE was present and showed that conventional methods to remove this bias were unsuccessful if exposure increased the probability of leaving work or of becoming ill—that is, when a time-varying confounder (leaving work or health status) was affected by prior exposure for some individuals. Arrighi and Hertz-Picciotto[7] used simulated data to illustrate (i) the failure of conventional methods to remove bias due to HWSE and (ii) the success of 'g-methods' proposed by Robins.[8–12]

Robins[8–12] developed g-methods specifically to eliminate HWSE and other biases caused by a time-varying confounder that is itself affected by prior exposure. One such method is g-estimation of structural accelerated failure time models. This method's first occupational application was published in 2012 by Chevrier *et al*,[13] who examined workplace exposure to straight metalworking fluids in the United Autoworkers—General Motors (UAW

To cite: Picciotto S, Brown DM, Chevrier J, *et al*. *Occup Environ Med* 2013;**70**:736–742.

—GM) cohort and compared the results to those obtained from conventional adjustment methods.

Occupational studies are often based on cohorts of active workers who are followed until they leave work.[14–17] By reanalysing the UAW—GM mortality study,[13] this pedagogical paper explores the potential for selection bias when outcome data past termination of employment are unavailable. Our objective is to provide a structural framework for understanding the HWSE that clarifies the prerequisites for using g-methods to avoid bias when follow-up ends at employment termination.

MATERIALS AND METHODS

Background: structure of the healthy worker survivor effect

As public health researchers, we are concerned with the total effect of workplace exposure, not simply the effect we can observe while workers are still employed. The effect measure we will attempt to estimate (our 'target parameter') is an overall health effect, quantified as the HR comparing a scenario in which all employees are exposed for 5 years with a scenario in which no one is ever exposed. Identifiability of this HR depends on several assumptions,[18] two of which are directly relevant to this study. Most epidemiologists are familiar with conditional exchangeability: the assumption that there is no unmeasured confounding. Positivity, or experimental treatment assignment, refers to the requirement that there be exposed and unexposed individuals in every non-empty stratum of the covariates.

We first consider conditional exchangeability as it relates to the HWSE. To elucidate the structure of this bias, we present directed acyclic graphs (DAGs) to make the relationships between variables visually explicit. In studies of cancer or chronic disease, epidemiologists are usually interested in the effects of cumulative exposure. In these DAGs, cumulative exposure is decomposed into a series of exposures over time in order to clarify the structure of the time-varying confounding. Figure 1A therefore shows a simplified DAG illustrating a time-varying exposure E_t, an outcome D and an unmeasured time-varying confounder: a health status U_t that, at each time t, affects the outcome and (indirectly) future exposure. (For simplicity, baseline covariates are suppressed and only two time points are included after baseline.) In addition, two mechanisms by which workers might reduce their exposure are present on pathways from health status to exposure: (i) taking time off work intermittently (leave of absence L_t=proportion of year t the individual did not work) and (ii) termination of employment (work status W_t=1 if the individual was actively employed in any part of year t, 0 if not). Health status U_t may cause the outcome D, and affects both W_t and L_t, but for now we assume that it affects future exposure only through these two pathways. Meanwhile, both intermittent time off work and work status influence future exposure, and they are also affected by prior exposure.

In a DAG, paths that do not include two arrowheads meeting at a variable (a 'collider' on the path) correspond to statistical association. Controlling for a variable on the path eliminates the association ('blocks the path'). If a path does include a collider, it is 'blocked' (does not represent a statistical association). However, controlling for the collider 'unblocks the path' so that a statistical association is artificially generated. For a detailed and accessible introduction to DAGs, see Hernán *et al*[19] or Greenland *et al*.[20]

In figure 1A, the paths $E_1 \leftarrow W_0 \leftarrow U_0 \rightarrow D$ and $E_1 \leftarrow L_0 \leftarrow U_0 \rightarrow D$ (and the corresponding paths one time point later) represent confounding.[19 20] Any occupational study whose assumptions are represented in this DAG must adjust for both work status and time off work: controlling for L_0 and W_0 will adjust for confounding of the effect of E_1 on D, and controlling for L_1 and W_1 will adjust for confounding of the effect of E_2 on D.

Since prior exposure also influences work status and intermittent time off work, W_t and L_t are exactly the type of variables that rendered conventional adjustment methods inadequate in the simulation studies mentioned above.[6 7] When using conventional methods to estimate the effect of duration of exposure or of cumulative exposure ($E_0+E_1+E_2$) on the outcome D, the required adjustment for these variables will create selection bias[19] (also called collider-stratification bias[21]). For example, W_0 is a collider on the path $E_0 \rightarrow W_0 \leftarrow U_0 \rightarrow D$ because the path includes two arrowheads colliding at W_0; adjusting for W_0 or stratifying by it therefore unblocks the path,[19] creating a spurious association between E_0 and D. Thus, our DAG illustrates that workers at different levels of exposure are neither exchangeable without stratification on W_t nor conditionally exchangeable within strata of W_t.

Background: adjusting for the healthy worker survivor effect

The DAG illustrates that when studying the effect of cumulative exposure, we must adjust for W_t and L_t, but that adjustment for

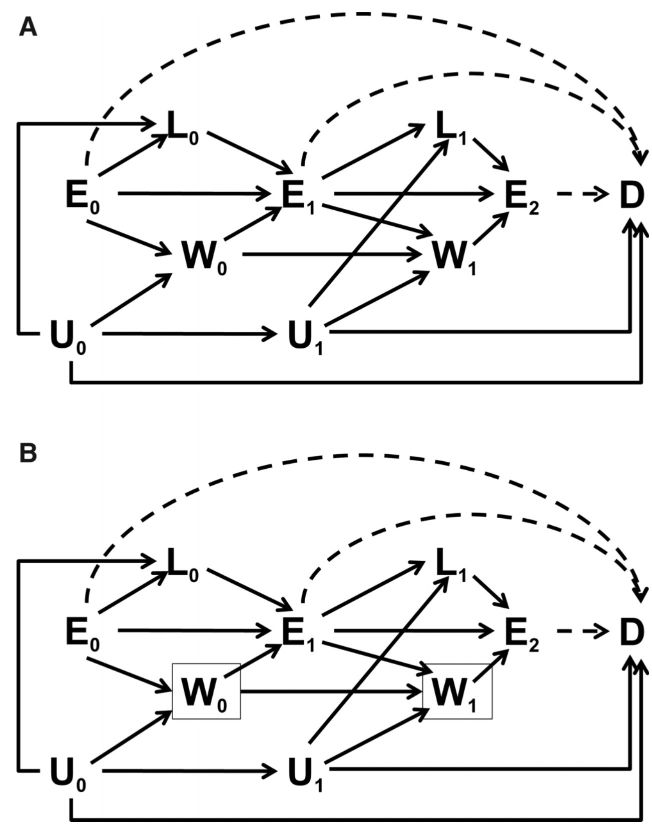

Figure 1 (A) Directed acyclic graph representing an occupational study with exposure E measured at times 0, 1 and 2, an outcome D at the end of follow-up, and an unmeasured health status variable U corresponding to the worker's underlying health at each time. The causal effect of interest is represented by the dashed arrows from exposure to outcome. The diagram also includes the measured variables L and W, representing (respectively) intermittent time off work and work status (actively employed or not) at each time. Diagram (B) represents the same study, restricted to actively employed person-time and outcomes occurring during active employment. The box around W at each time indicates the condition on active employment.

W_t and L_t without bias is impossible using conventional statistical analyses. This conclusion is supported by the simulation studies[6 7] and by Chevrier *et al*'s comparison of results using conventional methods and g-estimation.[13] Robins developed g-methods with occupational applications in mind;[8] they are necessary (but may not be sufficient) for proper adjustment for HWSE.[1 8 18 20]

We now consider how the positivity assumption limits the available methods to adjust for HWSE. The g-method that has been most widely adopted is inverse probability of treatment-weighted marginal structural models. Unfortunately, this method is prone to bias when the positivity assumption is violated, as it is in occupational studies if follow-up continues past termination of employment, when exposure cannot occur. Our target parameter refers to the entire follow-up, not just actively employed time, so inverse probability of treatment weighting is not appropriate for this application. G-estimation of a structural accelerated failure time model does not depend on the positivity assumption, though it does require extrapolation based on the model. We therefore apply this method and assume that the effect of exposure in the people who have already terminated employment would be the same as in the actively employed.

G-estimation of a structural accelerated failure time model has already been described in detail.[13 22] Briefly, the method requires two models: (i) a structural model for the counterfactual survival time if never exposed, which is a function of observed survival time, observed exposures and an unknown coefficient to be estimated and (ii) a model predicting exposure at each time. The exposure prediction model is where confounder adjustment takes place; all measured covariates are included in this model. By predicting exposure at each time on the basis of only the *prior* values of covariates rather than modelling the outcome conditional on exposure and covariate history, g-estimation disentangles the time-varying confounder relationships, avoiding the collider-stratification bias described above.

Under the (conditional exchangeability) assumption of no unmeasured confounders, the survival time if never exposed is statistically independent of observed exposure, within strata of the measured confounders. G-estimation refers to using optimisation methods to estimate the value of the unknown coefficient in the structural model (i) that achieves this statistical independence. Understanding the mechanics of the method is not required to follow the main points of this paper; interested readers should see Hernán *et al*[22] and Chevrier *et al*[13] for more details on the method and its application, including how to convert the estimated coefficient to a HR.

Under the assumptions of figure 1A, the target parameter (a HR comparing scenarios where the entire population is either exposed for 5 years or never exposed) can be estimated from the full follow-up dataset.[18] The previous application of g-estimation to the UAW—GM cohort examined the relationships of 5 years of exposure to straight metalworking fluids with the time to death from several different causes.[13] Censoring by competing risks or loss to follow-up was handled by inverse probability weighting. We assume that the DAG of figure 1A applies to this study.

The present paper expands on this work by examining the impact of restricting the analysis to actively employed person-time, yielding the dataset we would have obtained if the outcome data had been available only for active workers. Figure 1B, depicting the restricted dataset, is identical to figure 1A except for the boxes around W_0 and W_1, which represent conditioning on work status. Note that because restricting to actively employed person-time is the same as conditioning on employment status, the DAG includes unblocked non-causal paths from the exposure to the outcome through the colliders we have conditioned on: $E_0 \rightarrow W_0 \leftarrow U_0 \rightarrow D$ and $E_1 \rightarrow W_1 \leftarrow U_1 \rightarrow D$.[19 20] In other words, we have induced collider-stratification bias by observing outcomes only when workers are still actively employed. Exposure and outcome will be associated in the dataset restricted to active employment even under the null hypothesis of no effect of exposure on outcome.

Because work status is constant over the observations included in the analysis of the restricted dataset, it represents a source of selection bias rather than a confounder.[19] Therefore, additional measures are required to adjust for work status as a censoring variable. In order to demonstrate that this restriction to active employment time precludes correct estimation of our target parameter in this cohort, results are compared with those from an analysis of the full follow-up dataset. We use the results from the full follow-up dataset as a proxy for the 'truth'.

Since our target parameter refers to the whole UAW—GM population, including those who are no longer actively employed, we would like to interpret our estimate from the restricted dataset as applying to that whole population (ie, we wish to assume this estimate is equal to that from the full follow-up dataset). Either of two possible assumptions would permit us to do so: (1) overall, workers who terminate employment have the same risk of the outcome as those who stay at work, or (2) *within strata of the other measured covariates*, workers who do and do not terminate employment have the same risk of the outcome. However, DAG 1B contradicts both of these assumptions: there is a path from employment status W_t to D via the common cause U, which cannot be blocked because U is not measured. The best we can do is to make the second assumption, which is weaker but which requires a different DAG (figure 2), in which an additional time-varying variable H_t (*eg,* a health-related factor such as angina) is measured on the pathway from U_t to W_t.

We analysed data from the UAW—GM cohort to evaluate potential bias due to truncation at termination of employment in a study of the effects of metalworking fluids on all-cause and

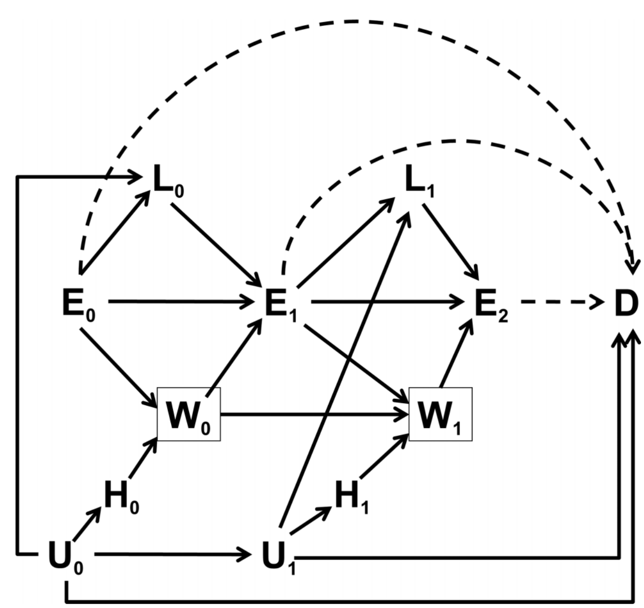

Figure 2 Directed acyclic graph representing an occupational study that includes additional measured covariates H allowing correct adjustment for censoring.

cause-specific mortality. The cohort consists of workers at three GM plants in Michigan, USA who were hired between 1938 and 1982 and who worked for at least 3 years. Mortality follow-up continued until the end of 1994 regardless of employment status. Exposure to metalworking fluids and time off work were determined by combining employment records with a job-exposure matrix developed in an extensive exposure assessment.[23]

We used the same g-estimation approach as Chevrier *et al*[13] except that we truncated person-time at termination of employment and censored survival times thereafter. Exposure in each year was binary (coded as 1 if ever exposed in that year, 0 otherwise). To adjust for confounding, probability of exposure in each year was predicted in a pooled logistic model including race, sex, age, age-squared, plant, exposures to metalworking fluids of three different types (straight, soluble, synthetic) in previous years and intermittent time off work. Note that the exposure prediction model only included the actively employed person-time for each individual, since those not at work could not be exposed. However, in the analysis of the full follow-up dataset, the structural model included survival time after the termination of employment. Censoring weights were calculated as in Hernán *et al*[22] based on probabilities estimated in a pooled logistic model among the people still at risk and still uncensored at each time, with the following predictors: race; sex; age at each time point and age at baseline; plant; both current and previous exposures to metalworking fluids of all three types; and intermittent time off work. The results from the weighted analysis correspond to a hypothetical pseudo-population in which no one is censored by loss to follow-up (less than 5%) or competing risks.

The present analysis includes 38 672 workers. The dataset restricted to actively employed person-time includes all individuals but excludes 36% of the total number of person-years from the full dataset. Due to loss of power in the restricted dataset, we considered only the three most common mortality outcomes from the original analysis: all causes, ischaemic heart disease and all cancers combined.

We used inverse probability weighting to adjust for censoring by leaving work, as well as for censoring by competing risks or loss to follow-up. Assuming (2) and DAG 2, our measured covariates are sufficient, and this analysis estimates results that would have been observed in a hypothetical pseudo-population in which nobody left work or got censored.[22] We used stabilised weights equal to the predicted marginal probability of remaining uncensored divided by the predicted conditional probability of remaining uncensored.[22] Conditional probabilities of being uncensored at each time were estimated from pooled logistic models run on those who were still at risk and uncensored, using the same exposure and covariate histories used in the original analysis[13] as predictors. These are essentially all the covariates available in this dataset; we assume that some of them play the role of H_t from figure 2 in the analysis. The conditional probabilities of being uncensored at each time were obtained from separate models for each type of censoring; the weights for leaving work and for competing risks and loss to follow-up were then multiplied together. Weights greater than 20 or less than 0.05 (<2%) were truncated to those maximum and minimum values, respectively.

In a sensitivity analysis, we omitted weights for leaving work, obtaining estimates of the effects among the active employees of an exposure intervention that allowed workers to terminate employment as they did in the observed data. This approach, which requires assumption (1) and a modified DAG 1B without U_t, instead of assumption (2) and DAG 2, is the censoring approach implicitly taken in most occupational studies where follow-up ends at employment termination. Under assumption (1), g-estimation without any adjustment for censoring by employment termination yields unbiased results.

CIs were obtained using 200 non-parametric bootstraps. All analyses were carried out in SAS V.9.3 (SAS Institute Inc, Cary, North Carolina, USA). The study involved reanalysis of existing data and was approved by the UC Berkeley Committee for the Protection of Human Subjects.

RESULTS

The cohort was mostly white (81%) and men (88%); workers remained at the company an average of 25 years. Table 1 summarises the distributions of baseline variables, the number of years each employee worked and the number of years exposed to straight metalworking fluids (among those who were ever exposed). Table 1 also shows the numbers of exposed and unexposed person-years during active employment for each type of metalworking fluid. Exposure to straight metalworking fluids occurred in 25% of the actively employed person-years included in the analysis.

Table 2 compares the two analysis datasets. Deaths of active employees tended to occur at an earlier age, and considerably more than half of the outcomes were censored in the restricted dataset.

The estimated HR for all-cause mortality, comparing the two exposure scenarios (whole population exposed to straight metalworking fluids for 5 years vs never exposed to straight metalworking fluids), was 1.09 (95% CI 1.07 to 1.11) in the full follow-up dataset, but 0.92 (95% CI 0.83 to 1.06) in the dataset restricted to active employment time only. HRs for ischaemic

Table 1 Descriptive characteristics and person-years of exposure to metalworking fluid types during active employment in the United Autoworkers—General Motors Cohort, Michigan, USA, 1941–1994

	N	Per cent
Race		
Black	7144	18.5
White	31 528	81.5
Sex		
Male	33 913	87.7
Female	4759	12.3
Plant		
1	9248	23.9
2	17 056	44.1
3	12 368	32.0
Age at baseline (mean (SD))	30.8	(9.1)
Years worked at end of follow-up (mean (SD))	25.1	(11.1)
Years exposed at end of follow-up (among the ever-exposed) (mean (SD))	16.1	(9.5)
Straight metalworking fluids (exposure of interest)		
Person-years unexposed	465189	74.7
Person-years exposed	157719	25.3
Soluble metalworking fluids		
Person-years unexposed	257593	41.4
Person-years exposed	365315	58.6
Synthetic metalworking fluids		
Person-years unexposed	521771	83.8
Person-years exposed	101137	16.2

Table 2 Distribution of demographic variables among person-years and among cases in the United Autoworkers—General Motors Cohort, Michigan, USA, 1941–1994, and in the dataset restricted to actively employed person-time only

	Full dataset	Restricted to active employment
Subjects	38672	38672
Person-years	969998	622908
% Male	89	89
% White	83	80
Age (mean)	45	40
All-cause mortality	9427	1617
% Male	93	96
% White	83	79
Age (mean)	62	49
Ischemic heart disease	2589	462
% Male	95	99
% White	89	89
Age (mean)	64	53
All cancers	2382	321
% Male	91	93
% White	83	79
Age (mean)	63	52

heart disease mortality and cancer mortality were similarly shifted downward in the censored dataset. Figure 3 shows these estimates for each outcome, juxtaposing results from the full follow-up dataset to those from the dataset restricted to active employment time only.

In the sensitivity analyses without censoring weights for employment termination, point estimates for ischaemic heart disease mortality and cancer mortality were very close to the corresponding HRs from the weighted analysis, albeit with wider CIs. For all-cause mortality, the sensitivity analysis without weights for employment termination yielded a HR estimate that was close to the null with a wide CI.

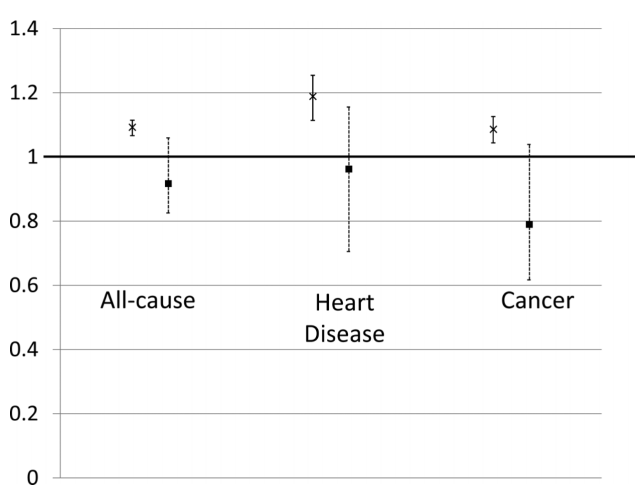

Figure 3 HRs comparing the counterfactual hazard if the whole United Autoworkers—General Motors cohort had been exposed (to straight metalworking fluids) for 5 years to the counterfactual hazard if the whole cohort had never been exposed, 1941–1994. HRs and 95% CIs for three outcomes for the full dataset (Xs with solid confidence bars) are compared with those for the dataset restricted to the actively employed person-time (squares with dashed confidence bars).

DISCUSSION

Whereas g-estimation in the full dataset indicated that 5 years of exposure to straight metalworking fluids was harmful for all three outcomes, the exposure appeared protective or neutral in the restricted dataset, with or without adjustment by weights for employment termination. We now revisit the reasons for these qualitative differences, pointing out several subtle but relevant aspects of the analysis.

One of the strengths of g-estimation is its ability to adjust for time-varying confounders that are affected by prior exposure. The three analyses of the UAW—GM cohort (full follow-up; restricted to active employment time with and without weights) adjusted for two different such variables: intermittent time off work and work status. In both the full follow-up dataset and the dataset restricted to actively employed person-time, g-estimation adjusted for intermittent time off work (L_t) by including it as a predictor in the model for exposure.

However, work status (W_t) required different handling in the full follow-up dataset and the dataset restricted to active employment time. Although g-estimation adjusted for intermittent time off work in both datasets, it adjusted for work status only in the full follow-up dataset. Recall that in the analysis of the full dataset, this confounder was not included in the exposure model, since those who were not actively employed could not be exposed. Instead, the exposure prediction model was conditional on not having left work previously. That is, even when analysing the full dataset, g-estimation adjusts for current work status by including only observations occurring during active employment. However, survival times are included in the structural model whether the outcome occurred before or after employment termination; collider-stratification bias is avoided by conditioning only on *current* work status in the exposure model.

By contrast, when we restricted to actively employed person-time, the *structural model* was applied only to those whose outcome occurred while they were still actively employed, thereby effectively conditioning on *future* work status. That is, being actively employed plays a different role in the two datasets. In the full dataset, work status is a time-varying confounder affected by prior exposure; g-estimation is specifically designed to adjust correctly for this type of confounding. However, in the restricted dataset, employment termination is a censoring variable that occurs *after* exposure. As our sensitivity analysis suggested, not adjusting for censoring by employment termination requires strong assumptions that are unlikely to be met. We believe, based on figure 1B, that this analysis suffers from collider-stratification bias because assumption (1) was violated: the analysis would only be unbiased if U were not a common cause of employment termination and the outcome.

Since employment termination is a *censoring* variable, not a confounder, in the restricted dataset, it requires additional adjustment using inverse probability weighting. In practice, this adjustment was unsuccessful: the effect estimates when censoring at employment termination, with or without weights, are all qualitatively different from the effect estimates in the full dataset, making exposure appear protective or neutral instead of harmful. Thus, a form of HWSE persists despite our use of g-methods, whether or not we attempt to adjust for employment termination; this probably means that at least one of our assumptions is not met.

The reason the adjustment did not work is that inverse probability weighting treats censoring as another 'exposure' on which to intervene. Weighting therefore requires that, like the exposed and the unexposed, the censored and the uncensored be conditionally exchangeable[18 19]: there should be no unmeasured 'confounding' between censoring and outcome. Within each stratum of the covariates included in the model predicting censoring, inverse probability weighting assigns weights to uncensored people so that they also represent censored people with the same characteristics. However, the adjustment is only valid if, within those strata, the censored and the uncensored have the same risk of experiencing the outcome. Unfortunately, in our study, those who left work permanently are probably more likely to be seriously ill than those who stayed at work, even within strata of our measured covariates. That is, the variables we measured did not really include H_t in figure 2; figure 1B is the DAG that really represents the cohort. Thus, there is unmeasured 'confounding' between censoring (leaving work) and the outcome, which cannot be fixed by weighting (or any method) without measuring additional variables.

The essential problem is that health status is an *unmeasured* common cause of employment status and the outcome. Just as we need measured variables in order to block paths from the exposure to the outcome via common causes, thereby controlling confounding, we also need measured variables to block paths from *censoring* to the outcome via common causes so that we can prevent collider-stratification bias. For example, in figure 2, let H_t represent angina or arrhythmia, which could be mechanisms by which underlying health status affects the decision to terminate employment. If these had been measured, then using them as predictors in the model for leaving work W_t would have allowed for better censoring adjustment.

The lack of such measured variables violates the assumption of conditional exchangeability of the censored and the uncensored (ie, the requirement that within strata of measured covariates, those who left work and those who did not have equal risk of the outcome). Our dataset does not provide other information about health status; we cannot block these paths because H_t is unmeasured. The results in the restricted dataset are therefore biased with respect to our target parameter. However, studies in which follow-up does end at employment termination may be more likely to measure health-related variables H_t.[14–16] Including H_t in the model predicting termination of employment for use in the inverse probability of censoring weights would mitigate the bias that occurred in our analysis.

Although we used the estimates from the full follow-up dataset as a proxy for the true value of our target parameter, those estimates may not be completely unbiased. For example, our assumption that the only two pathways by which health status might affect future exposure were by way of intermittent time off work or employment termination may be false. In reality, employees might reduce their exposure by changing jobs (without leaving the company or taking time off), or by using protective equipment. These pathways were assumed not to occur in our data. We also assumed that (i) there were no other unmeasured common causes of exposure and outcome or of loss to follow-up and outcome, (ii) all variables were measured without error, (iii) we had specified all models correctly, and (iv) duration of exposure, rather than cumulative exposure or exposure intensity, was the appropriate metric.

This application and the DAGs presented here illustrate the challenges of analysing occupational cohort data when follow-up ends at termination of employment. Although this is an artificial problem in the UAW—GM study, morbidity outcomes are censored at termination of employment in many occupational studies. In such cases, successful adjustment may be possible using inverse probability weighting if appropriate variables are included in the model for censoring by employment termination. For example, results from inverse probability weighted g-estimation using a well-chosen, measured health variable H_t that affects employment termination would not suffer from the substantial bias we found in this application.

The HWSE is complex: any mechanism by which workers reduce their exposures may cause bias in a conventional analysis. G-estimation, however, eliminates that bias unless the mechanism is also a censoring variable, as employment termination often is. In that case, g-estimation results will suffer from selection bias if appropriate health-related variables are not used to adjust for censoring in the analysis.

Contributors SP designed, implemented, and interpreted the analysis and wrote the article. DMB helped with the analysis design and interpretation, and performed additional analyses. JC performed preliminary data preparation and analyses and contributed to the writing. EAE designed the study, collected the original data, and supervised the analysis and the writing of the manuscript.

Funding This work was supported by the National Institute for Occupational Safety and Health at the Centers for Disease Control (grants R01OH009939 subcontract 28915360-48844B to E.A.E., S.P., and D.M.B., R01OH008927 to E.A.E., S.P., and J.C., R03OH010202 to S.P. and E.A.E., and R01OH010028 to E.A.E. and S.P.).

Competing interests None.

Ethics approval UC Berkeley Committee for the Protection of Human Subjects.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Eisen EA, Robins JM, Picciotto S. Healthy worker effect in occupational studies. In: El-Shaarawi A-H, Piegorsch W eds. *Encyclopedia of environmetrics*. Chichester, UK: John Wiley & Sons, 2012:1269–72.
2 Eisen EA. Healthy worker effect in morbidity studies. *Med Lav* 1995;86:125–38.
3 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.
4 Gilbert ES. Some confounding factors in the study of mortality and occupational exposures. *Am J Epidemiol* 1982;116:177–88.
5 Gilbert ES, Marks S. An analysis of the mortality of workers in a nuclear facility. *Radiat Res* 1979;79:122–48.
6 Steenland K, Deddens J, Salvan A, *et al*. Negative bias in exposure-response trends in occupational studies: modeling the healthy workers survivor effect. *Am J Epidemiol* 1996;143:202–10.
7 Arrighi HM, Hertz-Picciotto I. The evolving concept of the healthy worker survivor effect. *Epidemiology* 1994;5:189–96.
8 Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods—Application to control of the healthy worker survivor effect. *Mathematical Modelling* 1986;7:1393–512.
9 Robins JM, Tsiatis AA. Correcting for noncompliance in randomized trials using rank preserving structural failure time models. *Commun Stat Theory* 1991;20:2609–31.
10 Robins J. Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika* 1992;79:321–34.
11 Robins J, Tsiatis AA. Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika* 1992;79:311–19.
12 Robins JM. Structural nested failure time models. In: Andersen PK, Keiding N, eds. *Survival analysis*. Chichester, UK: John Wiley and Sons, 1998:4372–89.
13 Chevrier 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.
14 Harris C, Eisen EA, Goldberg R, *et al*. 1st place, PREMUS best paper competition: workplace and individual factors in wrist tendinosis among blue-collar workers—the San Francisco study. *Scand J Work Environ Health* 2011;37:85–98.
15 Eisen EA, Wegman DH, Louis TA. Effects of selection in a prospective study of forced expiratory volume in Vermont granite workers. *Am Rev Respir Dis* 1983;128:587–91.
16 Clougherty JE, Eisen EA, Slade MD, *et al*. Workplace status and risk of hypertension among hourly and salaried aluminum manufacturing employees. *Soc Sci Med* 2009;68:304–13.
17 Rabinowitz PM, Galusha D, McTague MF, *et al*. Tracking occupational hearing loss across global industries: a comparative analysis of metrics. *Noise Health* 2012;14:21–7.
18 Robins JM, Hernán MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G, eds. *Longitudinal data analysis*. New York: Chapman & Hall / CRC, 2009:553–99.

19 Hernán MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15:615–25.
20 Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;10:37–48.
21 Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology* 2003;14:300–6.
22 Hernán MA, Cole SR, Margolick J, *et al*. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiol Drug Saf* 2005;14:477–91.
23 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.

Healthy worker survivor bias: implications of truncating follow-up at employment termination

Sally Picciotto, Daniel M Brown, Jonathan Chevrier, et al.

Occup Environ Med 2013 70: 736-742 originally published online July 19, 2013
doi: 10.1136/oemed-2012-101332