How to identify the healthy worker survivor effect empirically and how to interpret results from published studies: the NIOSH ethylene oxide cohort as a case study

Authors: Sally Picciotto (1), Kaitlin Kelly-Reif (2), Ellen A. Eisen (1), Leslie T. Stayner (3), Sadie Costello (1)

Affiliations: 1. Environmental Health Sciences, University of California, Berkeley School of Public Health. 2. Division of Field Studies and Engineering, National Institute for Occupational Safety and Health. 3. Department of Epidemiology and Biostatistics, University of Illinois Chicago.

Acknowledgements: The authors thank Irva Hertz-Picciotto for helpful comments on an earlier version of the manuscript.

Funding: This work was funded by Intergovernmental Personnel Agreements provided by the National Institute for Occupational Safety and Health (#24IPA2411181 and # 24IPA2411182) and by an Interagency Agreement with the California Office of Environmental Health Hazard Assessment (#23-E0043).

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of NIOSH, the Centers for Disease Control and Prevention, or the California Office of Environmental Health Hazard Assessment.

Abstract

Modern causal methods are underutilized in occupational epidemiology, despite the development of robust methods to adequately control time-dependent confounding affected by prior exposure, the root of the healthy worker survivor effect. We demonstrate how to detect the healthy worker survivor effect empirically and explain how to interpret analyses that have not adjusted for it. For lymphohematopoietic cancer mortality and female breast cancer mortality, we performed pathway analyses assessing whether employment is a time-varying confounder affected by prior workplace exposure to ethylene oxide. These analyses ascertained whether the relevant causal relationships depicted in a directed acyclic graph were present. For both outcomes, workers employed longer were at lower risk. Workers exposed to higher levels of ethylene oxide were also more likely to leave work. Thus, employment is a time-varying confounder affected by prior exposure. The directions of these associations imply that healthy worker survivor effect is operating. Previously published estimates of health effects of workplace exposures to ethylene oxide on both lymphohematopoietic cancer mortality and female breast cancer mortality are underestimates of the true impacts. Applying these methods to other occupational cohorts can aid interpretations of analyses that have not adjusted for the healthy worker survivor effect.

The highest-quality epidemiologic data regarding human carcinogens frequently originate from occupational cohorts, exemplified by research on substances such as diesel exhaust,[1] benzene[2], and radon.[3] These cohorts provide important evidence of health effects even though they are observational studies, which are inherently vulnerable to selection bias, confounding, and other types of bias. In particular, occupational cohorts have substantially influenced our understanding of human carcinogens despite being susceptible to the healthy worker effect, which can result in estimates erroneously suggesting that an exposure is less harmful than it really is, or even that it is beneficial.[4–7] Over the past two decades, the healthy worker effect has become a well-documented methodological concern, with accepted study design and analytical methods developed to account for it.[8–14] The healthy worker effect acknowledges that (1) healthier individuals are more likely to seek employment and be hired initially (healthy hire effect), and (2) individuals who do not experience adverse health effects of occupational exposures tend to remain employed longer. Over the course of follow-up, the latter phenomenon produces time-dependent confounding by a mediator, the root of the healthy worker survivor effect.[4–7]

Despite scientific acceptance of methods to mitigate healthy worker effect and limited adoption of them,[9–30] most published occupational cohort studies do not apply these approaches. This omission places readers—and policymakers—in the challenging position of interpreting results that are biased toward the null (or beyond) because they are not adjusted for healthy worker effect. The situation poses two significant obstacles to environmental risk assessment and the regulation process: (1) Regulators may inadvertently select suboptimal action levels if the published literature and risk assessments are biased toward the null from healthy worker effect. (2) The evaluation of regulations is sometimes prolonged unnecessarily by misunderstandings of the impact of unaddressed healthy worker effect.

NIOSH Ethylene Oxide Cohort: A Detailed Examination

Consider the National Institute for Occupational Safety and Health (NIOSH) cohort established to investigate the health effects of ethylene oxide (EtO).[31–35] EtO is widely used to produce ethylene glycol for antifreeze and to sterilize medical equipment and spices, causing both workplace exposure and emissions into the ambient environment. Data derived from this cohort significantly contributed to the classification of EtO as a known human carcinogen by several authoritative agencies, including the National Toxicology Program Report on Carcinogens[36] and the International Agency for Research on Cancer,[37] which in turn motivated the United States Environmental Protection Agency to consider new regulations on emissions.[38–40]

Like many influential occupational cohorts, the NIOSH EtO cohort was rigorously designed to inform policy aimed at protecting individuals from harmful exposures.[31] The initial publication in 1991 presented standardized mortality ratios;[32] subsequent papers presented results of quantitative exposure-response associations between EtO and, specifically, lymphohematopoietic and breast cancer incidence and mortality, extended the follow-up period, and elaborated on the shape of the exposure-response curve.[33–35,41] The most recent publication from this cohort included 62 years of follow-up for female breast cancer mortality and thoroughly explored the shape of the exposure-response.[42]

Notably, none of the exposure-response papers from this cohort comprehensively assessed the presence of healthy worker survivor effect or addressed it in the analyses. However, some authors implicitly suggested the potential presence[42] or absence[34] of healthy worker survivor effect in the cohort based on theoretical considerations, and one author did a partial assessment.[43]

Because this cohort provides the strongest epidemiologic evidence for EtO carcinogenicity, its findings have been heavily featured in proposed regulations from the United States Environmental Protection Agency[40] and a draft update to an Inhalation Unit Risk factor from California's Office of Environmental Health Hazard Assessment.[44,45] Results from the NIOSH EtO cohort were also central to a report on EtO carcinogenicity[46] by the Texas Commission on Environmental Quality (TCEQ), recently reviewed by the National Academy of Sciences.[47] The TCEQ Report concluded that healthy worker effect was irrelevant in the case of EtO and female breast cancer risk, which they then used as justification to disregard the exposure-response analysis—an interpretation that was criticized by the National Academy of Sciences. Here we directly examine the empirical evidence for healthy worker survivor effect in the NIOSH EtO cohort for both lymphohematopoietic and breast cancer mortality.

The potential impact of healthy worker survivor effect on study results for EtO is currently relevant and of significant public health importance.[47] Using the current landscape of EtO regulation and the NIOSH EtO cohort as a case study, this paper aims to: (1) highlight key features of healthy worker survivor effect for readers of scientific literature, providing insights for interpreting results even in the potential presence of healthy worker survivor effect, and (2) determine whether a healthy worker survivor effect is operating within the NIOSH EtO cohort using a set of standard regressions to model pathways, providing a roadmap for similar analyses in other occupational cohorts.

Healthy Worker Survivor Effect

The healthy worker survivor effect makes health effects of occupational exposures difficult to study. This phenomenon occurs when the workers at highest risk of becoming ill leave employment due to their exposure, so that the workers who accumulate the most exposure are precisely those at lower risk of the outcome.[4–7] This sometimes leads to estimates suggesting that the exposure is beneficial, but even in less extreme examples, the healthy worker survivor effect results in underestimates of harmful effects. That is, an estimate of a positive association between exposure and outcome does not imply that the healthy worker survivor effect is absent; estimates can still be biased toward the null even if they do not cross over it.[5–8]

The healthy worker survivor effect arises due to a particular data structure in which employment is a time-varying confounder *and* is affected by prior exposure. By "time-varying confounder," we mean that employment status (which changes when a person terminates employment) is either a shared cause of exposure and outcome or is on a pathway connecting them through a shared cause—thus, it is related to risk of the health outcome. If employment status is not affected by prior exposure, then simple adjustment for employment prevents the healthy worker survivor effect; this scenario is not discussed further in this paper. However, if employment is affected by prior exposure, then collapsing exposures over time into a single measure (such as cumulative exposure) gives rise to a paradox whereby employment is a confounder that should be adjusted for, but since it depends on exposure, it is also potentially a mediator that should not be conditioned on when estimating total effects. Traditional regression methods are then guaranteed to produce a biased result regardless of whether they adjust for employment (including employment duration). Under certain assumptions, more complex statistical approaches called g-methods can disentangle exposure and employment over time and produce estimates that are not subject to this bias, as was demonstrated nearly 40 years ago.[4,5,8]

Many papers over the past decade or so have applied g-methods to occupational cohort studies.[9–30] Several have justified the approach by providing an assessment of the components of the healthy

worker survivor effect to show that the relationships in Figure 1 apply.[16,20,25,48,49] If this assessment determines that employment is *not* a time-varying confounder affected by prior exposure, then traditional regression analysis is justified. However, if it determines that employment *is* a time-varying confounder affected by prior exposure, then traditional regression analysis will yield biased estimates regardless of whether employment status (or its duration) is adjusted for.[5,8] Here we present an analysis to determine whether the healthy worker survivor effect is operating in the NIOSH EtO cohort by assessing whether employment in sterilizing facilities is a time-varying confounder affected by prior exposure to EtO in analyses of lymphohematopoietic or female breast cancers.

Methods

The cohort has been described previously.[31–34,42] Briefly, NIOSH assembled a cohort of workers exposed to EtO in 14 different sterilization facilities. Workers were originally followed up for mortality through 1987; follow-up was later extended to 1998[34] and most recently to 2021.[42] The original cohort included over 18,000 men and women with at least 3 months of exposure to EtO; however, 705 working at a plant lacking exposure estimates were therefore excluded from exposure-response analyses in earlier papers as well as here. Following the recent updated breast cancer analysis in this cohort,[42] we also excluded workers employed less than one year, who are likely to be systematically different from those who worked for at least one year, and we excluded the few who died before 1960, when outcome follow-up started. The study was reviewed and approved by the NIOSH Institutional Review Board. Informed consent was waived for this records-based study.

Unpacking exposures and other covariates over time is key to understanding the healthy worker survivor effect. Some familiarity with directed acyclic graphs is needed to understand the following explanation. Figure 1 is a simplified, incomplete causal directed acyclic graph[50,51] illustrating certain key variables at only three time points (the actual study covers multiple decades) and highlighting which causal relationships are relevant to determining whether the healthy worker survivor effect is present. Since only those employed are exposed at work (Figure 1B), employment status affects exposure, a relationship that does not need to be assessed empirically. The remaining criteria for the healthy worker survivor effect, discussed in more detail below, are that employment should (1) either affect the outcome or share a cause with it (Figure 1C or 1D, respectively) and (2) be affected by prior exposure (Figure 1E); adjusted regression models can help determine whether these criteria hold. The directions of the resulting associations matter; see below. The seminal paper on pathway analyses to evaluate the presence of the healthy worker survivor effect[48] and a prior simulation study[52] emphasize that the magnitude of the point estimate is a better indicator for this purpose than its statistical significance.

If both criteria hold, then exposure-response regression models *should* adjust for employment status in order to remove confounding of the effect of exposure at time t; however, they simultaneously *should not* condition on it because doing so could both remove part of the effect of exposure at time t-1 (and at earlier time points not depicted) and introduce collider bias due to the path from exposure at time t-1 (or earlier time points) through employment status at time t and the unmeasured health status at time t to the outcome.[50] Thus, exposure-response regressions using cumulative exposure that do not adjust for employment are subject to confounding, while models that condition on employment status or duration of employment are also subject to bias. If the healthy worker survivor effect is present, then g-methods are required to estimate the effect of exposure on the outcome without bias, since (unlike traditional regression) they can adjust for employment without conditioning on it.[8]

A prior paper on this cohort with follow-up ending in 1998 included a partial assessment of the healthy worker survivor effect, evaluating only one of the two criteria and providing evidence that employment was affected by prior exposure.[43] For completeness, and because follow-up has been extended to 2021, we assess both criteria here.

Criterion 1

Does employment affect or share a cause with the outcome, with workers who remain at work longer experiencing lower risk of the outcome? An association between the outcome and a function of employment status up to that point, conditional on other known confounders of the exposure-outcome relationship, would indicate that employment does affect or share a cause with the outcome. Previous papers presenting a pathway analysis to assess the presence of the healthy worker survivor effect did not discuss the direction of the association.[16,20,25,48,49] Employment is a time-varying confounder if this association is present, regardless of its direction; however, the direction of the association matters for identifying the healthy worker survivor effect. If a healthy worker survivor effect is operating, then the healthiest workers with lowest risk would have been able to remain employed *longer*. We thus addressed this question by regressing each outcome on employment duration. Since those at higher risk might also have *left work at younger ages*, another approach (not presented here) would be to regress each outcome on age at employment termination.[49]

Specifically, we fit Cox proportional hazards models for all lymphohematopoietic cancer mortality, using age as the time scale, with time-varying employment duration (as a linear predictor or in categories) as the main predictor, using person-time from the later of cohort entry or 1960 to end of follow-up. In these analyses we adjusted for race, plant, sex, calendar year, year of cohort entry, and cumulative exposure up to the previous year, which were hypothesized to potentially affect both employment status and cancer mortality. We then repeated these analyses for breast cancer mortality among women, omitting adjustment for sex. Magnitudes and signs of the estimates from these analyses tell us whether Criterion 1 is satisfied.

Criterion 2

Does higher exposure in the prior time period increase the probability of employment termination? EtO could cause some workers to terminate employment earlier than they otherwise would have, thus accumulating less exposure than the "healthy" workers who tolerated exposure better. (This is especially likely because EtO is not just a carcinogen but also an irritant.[53] However, non-irritant exposures can also satisfy this condition.) To assess this, we used a similar approach to ones used by Neophytou et al.[16] and Bertke et al.,[25] regressing employment termination in each year on the exposure accrued in the previous year, ending follow-up at employment termination. We used the *previous* year's exposure in order to avoid reverse-causation bias: if someone leaves work mid-year, they have less time to accumulate exposure in that year. Other options would be to consider average exposure intensity over that previous year or peak exposure in that previous year. Analyses could also be run on a finer time scale if appropriate for the exposure in question and if the data are available.

Those familiar with the occupational or environmental epidemiology literatures may believe that cumulative exposure is the preferred biologic exposure metric of interest and should therefore be used in the assessment of Criterion 2. Indeed, many assessments of Criterion 2 in the literature have assessed whether cumulative exposure affects employment termination.[20,25,43,48,49] However, cumulative exposure

is heavily influenced by duration of employment (itself a function of employment status, a key variable in the confounding pathway [Figure 1]), potentially leading to a similar healthy worker survivor effect: those who accumulate the most exposure are the ones who work longest and are *less* likely to leave work. Our goal is not to establish that the exposure metric used in the exposure-response model is associated with leaving work, but rather to determine whether the causal relationship depicted in Figure 1 exists. Departing from common assumptions about the exposure metric in the pathway analysis having to match the exposure metric in the exposure-response analysis, we are guided by the causal diagram in Figure 1, which specifically breaks down cumulative exposure into annual (or daily) exposures over time in order to illustrate the components of the healthy worker survivor effect that we are assessing. Note that breaking down cumulative exposure into separate exposures over time is the only way to illustrate (or assess, or adjust for) the healthy worker survivor effect.[5,8] Although cumulative exposure is also time-varying, only non-cumulative exposure can disentangle the relationships between exposure and employment over time.[25] We therefore attempted to isolate the effect of E_{t-1}, exposure in the previous year, on employment at time t, as shown in Figure 1E.

For this criterion, we ran Cox proportional hazards models for employment termination, using age as the time scale, with exposure in the prior year (time-varying, categorized in quartiles) as the main predictor, using person-time from the later of cohort entry or 1960 to employment termination. We adjusted for race, plant, sex, calendar year, year first employed, and earlier exposures (lags 2-5), since these could affect both exposure and employment status. This criterion does not depend on the outcome; however, since the breast cancer analysis is only in women, we assessed this criterion separately for the whole cohort and for the female subcohort (omitting adjustment for sex). Magnitudes and signs of the estimates from these analyses tell us whether Criterion 2 is satisfied, assuming consistency (*i.e.*, that exposure is well-defined), conditional exchangeability (*i.e.*, that we adjusted for all confounders of the relationship between employment termination and exposure in the previous year), and positivity (*i.e.*, that probabilities of both leaving work and remaining at work are nonzero for all possible values of exposure). In sensitivity analyses, we included person-time before 1960 (10,358 person-years overall; 5611 among the women): some workers (hired starting in 1918) were "at risk" of leaving employment before 1960, when follow-up began for the outcomes.

Results

Table 1 describes the cohort for the lymphohematopoietic cancer pathway analyses, and Table 2 describes the cohort restricted to women for the breast cancer pathway analyses. There were more women than men in the cohort overall, which was predominantly white; over 75% were born prior to 1950.

Table 3 presents results from the analyses for Criterion 1, concerning whether employment causes or shares a cause with the outcomes. Each additional year of employment was associated with slightly lower hazard of both outcomes. The categorical analyses also suggested a monotonic, negative relationship between employment duration and disease, which (at least for breast cancer) appeared to be non-linear.

Table 4 presents results from the analyses for Criterion 2, assessing whether employment is affected by prior exposure. Every category of exposure above the lowest quartile was associated with elevated hazard of employment termination, though this association appeared non-monotonic. Estimates were

slightly stronger among women. In the sensitivity analyses for Criterion 2, which included person-time from hire until employment termination (rather than starting from the later of 1960 or hire year), estimates and confidence limits differed from those presented here by less than 0.1.

Discussion

The fact that EtO is a known irritant[53] makes a healthy worker survivor effect particularly likely, and these pathway analyses provide strong evidence in support of that hypothesis, both for lymphohematopoietic cancers in all employees and for breast cancer among women. In the analyses assessing Criterion 1, employment duration was associated with both outcomes, showing that employment is a time-varying confounder; longer employment being associated with lower risk means that Criterion 1 is satisfied. Furthermore, employment is affected by prior exposure, as seen in the analyses assessing Criterion 2. These associations were in the direction consistent with a healthy worker survivor effect: those employed longer were at lower risk of the mortality outcomes under study; higher exposure accrued in a year was associated with greater probability of then leaving employment. Thus, employment is a time-varying confounder affected by prior exposure, with healthier workers likely to accrue more exposure, so there is downward bias (toward or even potentially across the null) in any estimates obtained from traditional analyses of EtO and mortality from either lymphohematopoietic cancers or breast cancer.

We estimated strong associations between duration of employment and both outcomes, and between exposure in a given year and employment termination in the following year. The association between duration of employment and lymphohematopoietic cancer mortality was not significant at the 0.05 level. However, as noted previously, the magnitude of the point estimate is a better indicator for this purpose than its statistical significance.[48,52] Indeed, assessment of confounding in observational studies does not generally rely on statistical significance; consider the classic change-in-point-estimate criterion as a case in point. Even if not statistically significant, strong associations along the component pathways can still produce substantial bias in the main exposure-response estimate. Furthermore, for specifically identifying the healthy worker survivor effect, the directions of the associations along the component pathways also matter.

Note that the presence of a healthy worker survivor effect does not provide clear evidence of the exact correct shape of the exposure-response curve, other than to lend credibility to curves that have a sharp rise followed by a flattening over higher levels of exposure.[54] While the steep slope at lower exposure levels is less affected by the healthy worker survivor effect, the highest exposures will only be accumulated by the least susceptible "healthiest" workers, so we expect the slope to be less steep at those higher exposures if the analysis does not adjust for the healthy worker survivor effect. Therefore, a simple linear term over the entire exposure distribution is likely to be a particularly poor fit for an occupational exposure-response curve using traditional regression, since the biased attenuation at the high end of exposure will artificially reduce the slope of the linear term; a two-piece linear spline or a non-linear model may fit the data better and highlight the steep slope at lower exposures (where the healthy worker survivor effect is less of a problem). Note, however, that these alternative models still do not solve the problem of the healthy worker survivor effect, particularly at higher cumulative exposures.

The presence of the healthy worker survivor effect also does not indicate the degree to which the main results will be biased. No methodology exists for adjusting the primary estimates to account for healthy worker survivor effect based on the results from the component pathway analysis. What *is* known is the

direction of the bias. When the associations are in the directions we observed in this cohort, the bias will correspond to a healthy worker effect: the healthiest workers experience higher cumulative exposure, so regression estimates will underestimate the risks associated with exposure.

Prior studies[32–34,42] have found elevated mortality from these cancers in association with EtO exposure. In the 1998 follow-up,[34] odds ratios for lymphohematopoietic cancer mortality for men were over 2.0 in all quartiles of unlagged cumulative exposure above the lowest; with a 15-year lag the trend was monotonic. For women, the odds ratio was elevated for the second quartile of unlagged cumulative exposure, but not for the higher quartiles; lagged results were not presented. However, given the results from this pathway analysis, the true effects in higher exposure groups are almost certainly stronger than previously estimated. Since the evidence presented here suggests that exposure was a strong influence on leaving work among women, the weak associations seen in previous publications should not be taken as proof that EtO exposure does not increase the risk of lymphohematopoietic cancer mortality in women; a healthy worker survivor effect may have obscured the true association. For breast cancer mortality, the 1998 follow-up estimated monotonically increasing odds ratios in all quartiles of non-zero cumulative exposure with a 20-year lag, reaching 3.13 (1.42, 6.92) in the highest category. When follow-up was extended to 2021, these results were confirmed, with elevated rates for all categories of non-zero cumulative exposure whether or not the analysis adjusted for duration of employment. Our findings provide evidence that even these positive findings are underestimates, which means that the effects of EtO on mortality from lymphohematopoietic and breast cancers are *at least as harmful* as the published papers report.

Some recent analyses in this cohort[42,43] have attempted to adjust for healthy worker survivor effect by adjusting for employment duration; however, as discussed above, this adjustment does not eliminate healthy worker survivor effect.[4–6,8] To fully account for healthy worker survivor effect, estimates of the potential impacts on mortality from these cancers if exposures to EtO were reduced could be obtained by analyzing the cohort data using g-computation, which can adjust for employment without conditioning on it, thus avoiding bias from the healthy worker survivor effect.[8] For now, this pathway analysis (a) serves to verify that the healthy worker survivor effect is operating in this cohort (and therefore that the true effects of EtO on these outcomes are stronger than previously estimated) and (b) represents a roadmap for assessing its presence in other cohorts, even if applying g-methods to mitigate the problem is not currently possible. Certainly, formal analyses of the components of healthy worker survivor effect in occupational cohort exposure-response papers would allow for evidence-based discussions of the presence of the healthy worker survivor effect and any potential attenuating bias.

References

1. Attfield MD, Schleiff PL, Lubin JH, Blair A, Stewart PA, Vermeulen R, Coble JB, Silverman DT. The diesel exhaust in miners study: A cohort mortality study with emphasis on lung cancer. *J Natl Cancer Inst*. 2012;104(11):869-883. doi:10.1093/jnci/djs035

2. Infante PF, Wagoner JK, Rinsky RA, Young RJ. LEUKÆMIA IN BENZENE WORKERS. *Lancet*. 1977;310(8028):76-78. doi:10.1016/S0140-6736(77)90074-5

3. Roscoe RJ, Deddens JA, Salvan A, Schnorr TM. Mortality among Navajo uranium miners. *Am J Public Health*. 1995;85(4):535-540. doi:10.2105/AJPH.85.4.535

4. Arrighi HM, Hertz-Picciotto I. The Evolving Concept of the Healthy Worker Survivor Effect.

Epidemiology. 1994;5(2):189-196.

5. Buckley JP, Keil AP, McGrath LJ, Edwards JK. Evolving Methods for Inference in the Presence of Healthy Worker Survivor Bias. *Epidemiology*. 2015;26(2):204-212. doi:10.1097/EDE.0000000000000217

6. Picciotto S, Hertz-Picciotto I. Commentary: Healthy Worker Survivor Bias: A Still-Evolving Concept. *Epidemiology*. 2015;26(2):213-215. doi:10.1097/EDE.0000000000000023300001648-201503000-00013 [pii]

7. Brown DM, Picciotto S, Costello S, Neophytou AM, Izano MA, Ferguson JM, Eisen EA. The Healthy Worker Survivor Effect: Target Parameters and Target Populations. *Curr Environ Heal reports*. 2017;4(3):364-372. doi:10.1007/S40572-017-0156-X

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. *Math Model*. 1986;7:1393-1512.

9. 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(2):212-219. doi:10.1097/EDE.0b013e318245fc0600001648-201203000-00006 [pii]

10. Naimi AI, Richardson DB, Cole SR. Causal inference in occupational epidemiology: accounting for the healthy worker effect by using structural nested models. *Am J Epidemiol*. 2013;178(12):1681-1686. doi:10.1093/aje/kwt215kwt215 [pii]

11. Cole SR, Richardson DB, Chu H, Naimi AI. Analysis of occupational asbestos exposure and lung cancer mortality using the g formula. *Am J Epidemiol*. 2013;177(9):989-996. doi:10.1093/aje/kws343kws343 [pii]

12. Naimi AI, Cole SR, Hudgens MG, Richardson DB. Estimating the effect of cumulative occupational asbestos exposure on time to lung cancer mortality: using structural nested failure-time models to account for healthy-worker survivor bias. *Epidemiology*. 2014;25(2):246-254. doi:10.1097/EDE.0000000000000004500001648-201403000-00015 [pii]

13. Neophytou AM, Costello S, Picciotto S, Noth EM, Liu S, Lutzker L, Balmes JR, Hammond K, Cullen MR, Eisen EA. Accelerated lung function decline in an aluminium manufacturing industry cohort exposed to pm 2.5: An application of the parametric g-formula. *Occup Environ Med*. 2019;76(12):888-894. doi:10.1136/oemed-2019-105908

14. Picciotto S, Chevrier J, Balmes J, Eisen EA. Hypothetical Interventions to Limit Metalworking Fluid Exposures and Their Effects on COPD Mortality: G-Estimation Within a Public Health Framework. *Epidemiology*. 2014;25(3):436-443. doi:10.1097/EDE.0000000000000082

15. Picciotto S, Peters A, Eisen EA. Hypothetical exposure limits for oil-based metalworking fluids and cardiovascular mortality in a cohort of autoworkers: structural accelerated failure time models in a public health framework. *Am J Epidemiol*. 2015;181(8):563-570. doi:10.1093/aje/kwu484kwu484 [pii]

16. Neophytou AM, Picciotto S, Costello S, Eisen EA. Occupational Diesel Exposure, Duration of Employment, and Lung Cancer: An Application of the Parametric G-Formula. *Epidemiology*. 2016;27(1):21-28. doi:10.1097/EDE.0000000000000389

17. Picciotto S, Ljungman PL, Eisen EA. Straight Metalworking Fluids and All-Cause and Cardiovascular Mortality Analyzed by Using G-Estimation of an Accelerated Failure Time Model With Quantitative Exposure: Methods and Interpretations. *Am J Epidemiol*. 2016;183(7):680-688. doi:10.1093/aje/kwv232kwv232 [pii]

18. Keil AP, Richardson DB. Reassessing the Link between Airborne Arsenic Exposure among Anaconda Copper Smelter Workers and Multiple Causes of Death Using the Parametric g-Formula. *Env Heal Perspect*. 2017;125(4):608-614. doi:10.1289/EHP438EHP438 [pii]

19. Neophytou AM, Picciotto S, Brown DM, Gallagher LE, Checkoway H, Eisen EA, Costello S. Estimating counterfactual risk under hypothetical interventions in the presence of competing events: Crystalline silica exposure and mortality from 2 causes of death. *Am J Epidemiol*. 2018;187(9):1942-1950. doi:10.1093/aje/kwy077

20. Keil AP, Richardson DB, Westreich D, Steenland K. Estimating the impact of changes to occupational standards for silica exposure on lung cancer mortality. *Epidemiology*. 2018;29(5):658-665. doi:10.1097/EDE.0000000000000867

21. Picciotto S, Neophytou AM, Brown DM, Checkoway H, Eisen EA, Costello S. Occupational silica exposure and mortality from lung cancer and nonmalignant respiratory disease. *Environ Epidemiol*. 2018;2(3):e029. doi:10.1097/ee9.0000000000000029

22. Garcia E, Picciotto S, Neophytou AM, Bradshaw PT, Balmes JR, Eisen EA. Lung cancer mortality and exposure to synthetic metalworking fluid and biocides: Controlling for the healthy worker survivor effect. *Occup Environ Med*. 2018;75(10):730-735. doi:10.1136/oemed-2017-104812

23. Izano MA, Sofrygin OA, Picciotto S, Bradshaw PT, Eisen EA. Metalworking Fluids and Colon Cancer Risk. *Environ Epidemiol*. 2019;3(1):e035. doi:10.1097/ee9.0000000000000035

24. Neophytou AM, Costello S, Picciotto S, Brown DM, Attfield MD, Blair A, Lubin JH, Stewart PA, Vermeulen R, Silverman DT, Eisen EA. Diesel Exhaust, Respirable Dust, and Ischemic Heart Disease: An Application of the Parametric g-formula. *Epidemiology*. 2019;30(2):177-185. doi:10.1097/EDE.0000000000000954

25. Bertke SJ, Keil AP, Daniels RD. Lung Cancer Mortality and Styrene Exposure in the Reinforced-Plastics Boatbuilding Industry: Evaluation of Healthy Worker Survivor Bias. *Am J Epidemiol*. 2021;190(9):1784-1792. doi:10.1093/aje/kwab108

26. Dumas O, Le Moual N, Siroux V, Heederik D, Garcia-Aymerich J, Varraso R, Kauffmann F, Basagana X. Work related asthma. A causal analysis controlling the healthy worker effect. *Occup Env Med*. 2013;70(9):603-610. doi:10.1136/oemed-2013-101362oemed-2013-101362 [pii]

27. Neophytou AM, Picciotto S, Hart JE, Garshick E, Eisen EA, Laden F. A structural approach to address the healthy-worker survivor effect in occupational cohorts: an application in the trucking industry cohort. *Occup Env Med*. 2014;71(6):442-447. doi:10.1136/oemed-2013-102017oemed-2013-102017 [pii]

28. Neophytou AM, Costello S, Brown DM, Picciotto S, Noth EM, Hammond SK, Cullen MR, Eisen EA. Marginal Structural Models in Occupational Epidemiology: Application in a Study of Ischemic Heart Disease Incidence and PM2.5 in the US Aluminum Industry. *Am J Epidemiol*. 2014;180(6):608-615. doi:10.1093/aje/kwu175kwu175 [pii]

29. Costello S, Picciotto S, Rehkopf DH, Eisen EA. Social disparities in heart disease risk and survivor

bias among autoworkers: an examination based on survival models and g-estimation. *Occup Env Med*. 2015;72(2):138-144. doi:10.1136/oemed-2014-102168oemed-2014-102168 [pii]

30. Keil AP, Richardson DB, Troester MA. Healthy worker survivor bias in the Colorado Plateau uranium miners cohort. *Am J Epidemiol*. 2015;181(10):762-770. doi:10.1093/aje/kwu348kwu348 [pii]

31. Stayner L, Steenland K. Fine Wines and Cohorts Take Time: The History of a Cohort Study of Workers Exposed to Ethylene Oxide. In: Dworkin MS, ed. *Cases in Field Epidemiology: A Global Perspective*. Jones & Bartlett Learning; 2011:373-388.

32. Steenland K, Stayner L, Greife A, Halperin W, Hayes R, Hornung R, Nowlin S. Mortality among Workers Exposed to Ethylene Oxide. *N Engl J Med*. 1991;324(20):1402-1407. doi:10.1056/NEJM199105163242004,

33. Stayner L, Steenland K, Greife A, Homung R, Hayes RB, Nowlin S, Morawetz J, Ringenburg V, Elliot L, Halperin W. Exposure-response analysis of cancer mortality in a cohort of workers exposed to ethylene oxide. *Am J Epidemiol*. 1993;138(10):787-798. doi:10.1093/OXFORDJOURNALS.AJE.A116782,

34. Steenland K, Stayner L, Deddens J. Mortality analyses in a cohort of 18 235 ethylene oxide exposed workers: follow up extended from 1987 to 1998. *Occup Env Med*. 2004;61(1):2-7.

35. Steenland K, Whelan E, Deddens J, Stayner L, Ward E. Ethylene oxide and breast cancer incidence in a cohort study of 7576 women (United States). *Cancer Causes Control*. 2003;14(6):531-539. doi:10.1023/A:1024891529592

36. *15th Report on Carcinogens*. National Toxicology Program; 2021. doi:10.22427/NTP-OTHER-1003

37. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 97. 1,3-Butadiene, Ethylene Oxide and Vinyl Halides (Vinyl Fluoride, Vinyl Chloride and Vinyl Bromide).* Vol 97.; 2008.

38. Federal Register :: National Emission Standards for Hazardous Air Pollutants: Ethylene Oxide Emissions Standards for Sterilization Facilities Residual Risk and Technology Review. Accessed October 17, 2025. https://www.federalregister.gov/documents/2024/04/05/2024-05905/national-emission-standards-for-hazardous-air-pollutants-ethylene-oxide-emissions-standards-for

39. USEPA. Final Amendments to Strengthen Air Toxics Standards for Ethylene Oxide Commercial Sterilizers. Published 2024. https://www.epa.gov/hazardous-air-pollutants-ethylene-oxide/final-amendments-strengthen-air-toxics-standards-ethylene

40. EnvironmentalProtectionAgency. *New Source Performance Standards for the Synthetic Organic Chemical Manufacturing Industry and Natural Emission Standards for Hazardous Air Pollutants for the Synthetic Organic Chemical Manufacturing Industry and Group I & II Polymers and Resins Industry*. 2024-07002; 2024:42932-43297. https://www.federalregister.gov/documents/2024/05/16/2024-07002/new-source-performance-standards-for-the-synthetic-organic-chemical-manufacturing-industry-and

41. Steenland K, Seals R, Klein M, Jinot J, Kahn HD. Risk estimation with epidemiologic data when response attenuates at high-exposure levels. *Environ Health Perspect*. 2011;119(6):831-837. doi:10.1289/ehp.1002521

42. Kelly-Reif K, Bertke SJ, Stayner L, Steenland K. Exposure to ethylene oxide and relative rates of female breast cancer mortality: 62 years of follow-up in a large US occupational cohort. *Environ Health Perspect*. Published online April 2025. doi:10.1289/EHP15566

43. Park RM. Associations between exposure to ethylene oxide, job termination, and cause-specific mortality risk. *Am J Ind Med*. 2020;63(7):577-588. doi:10.1002/AJIM.23115,

44. OEHHA. *Ethylene Oxide Cancer Inhalation Unit Risk Factor Technical Support Document. April 2023 Draft.*; 2023. https://oehha.ca.gov/sites/default/files/media/downloads/crnr/etocanceriurdraft040723.pdf

45. No Significant Risk Levels for Ethylene Oxide. Published 2024. Accessed July 28, 2025. https://oehha.ca.gov/proposition-65/crnr/no-significant-risk-levels-ethylene-oxide

46. Haney Jr. JT, Lange SS, Jenkins A, Myers JL. *Ethylene Oxide Carcinogenic Dose-Response Assessment*.; 2020. https://www.tceq.texas.gov/downloads/toxicology/dsd/final/eto.pdf

47. *Review of Texas Commission on Environmental Quality's Ethylene Oxide Development Support Document*. National Academies Press; 2025. doi:10.17226/28592

48. Naimi AI, Cole SR, Hudgens MG, Brookhart MA, Richardson DB. Assessing the component associations of the healthy worker survivor bias: occupational asbestos exposure and lung cancer mortality. *Ann Epidemiol*. 2013;23(6):334-341. doi:10.1016/j.annepidem.2013.03.013S1047-2797(13)00089-6 [pii]

49. Garcia E, Picciotto S, Costello S, Bradshaw PT, Eisen EA. Assessment of the healthy worker survivor effect in cancer studies of the United Autoworkers-General Motors cohort. *Occup Env Med*. 2017;74(4):294-300. doi:10.1136/oemed-2016-104038oemed-2016-104038 [pii]

50. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37-48. doi:00001648-199901000-00005 [pii]

51. Digitale JC, Martin JN, Glymour MM. Tutorial on directed acyclic graphs. *J Clin Epidemiol*. 2022;142:264-267. doi:10.1016/j.jclinepi.2021.08.001

52. Naimi AI, Cole SR, Westreich DJ, Richardson DB. A comparison of methods to estimate the hazard ratio under conditions of time-varying confounding and nonpositivity. *Epidemiology*. 2011;22(5):718-723. doi:10.1097/EDE.0b013e31822549e8

53. ATSDR. *Toxicological Profile for Ethylene Oxide (TP 137)*.; 1990. https://www.atsdr.cdc.gov/toxprofiles/tp137.pdf

54. Stayner L, Steenland K, Dosemeci M, Hertz-Picciotto I. Attenuation of exposure-response curves in occupational cohort studies at high exposure levels. *Scand J Work Environ Heal*. 2003;29(4):317-324. doi:10.5271/sjweh.737

Table 1. Descriptive characteristics of cohort of U.S. sterilization facility workers exposed to ethylene oxide, 1960-2021.

	N	%	mean	median	IQR	
Number of workers	13215	100				
Year of birth			1938	1941	1928	1949
Sex						
Male	5666	43				
Female	7549	57				
Race						
White	10529	80				
Non-white	2686	20				
Employment duration			10.6	7.4	2.7	17.0
Exposure duration			6.6	3.8	1.6	9.8
Cumulative exposure to EtO (ppm-days)			12503	3761	960	12567
Average annual accrued exposure when exposed (ppm-days)			1850	977	359	2451
Deceased	7124	54				
Lymphohematopoietic cancer death	167	1				
Age at lymphohematopoietic cancer death			70	72	63	79
Age at first employment			31	28	22	38
Age at end of follow-up			73	74	66	81
Person-years of follow-up			41	44	36	50

Table 2. Descriptive characteristics of women in the cohort of U.S. sterilization facility workers exposed to ethylene oxide, 1960-2021.

	N	%	mean	median	IQR	
Number of workers	7549	100				
Year of birth			1937	1939	1927	1948
Race						
White	6120	81				
Non-white	1429	19				
Employment duration			10.7	7.4	2.7	17.4
Exposure duration			6.4	3.5	1.6	9.6
Cumulative exposure to EtO (ppm-days)			8503	3056	865	9712
Average annual accrued exposure when exposed (ppm-days)			1353	885	369	1869
Deceased	4010	53				
Breast cancer death	181	2				
Age at breast cancer death			65	64	55	75
Age at first employment			31	30	22	39
Age at end of follow-up			75	76	68	83
Person-years of follow-up			42	45	37	50

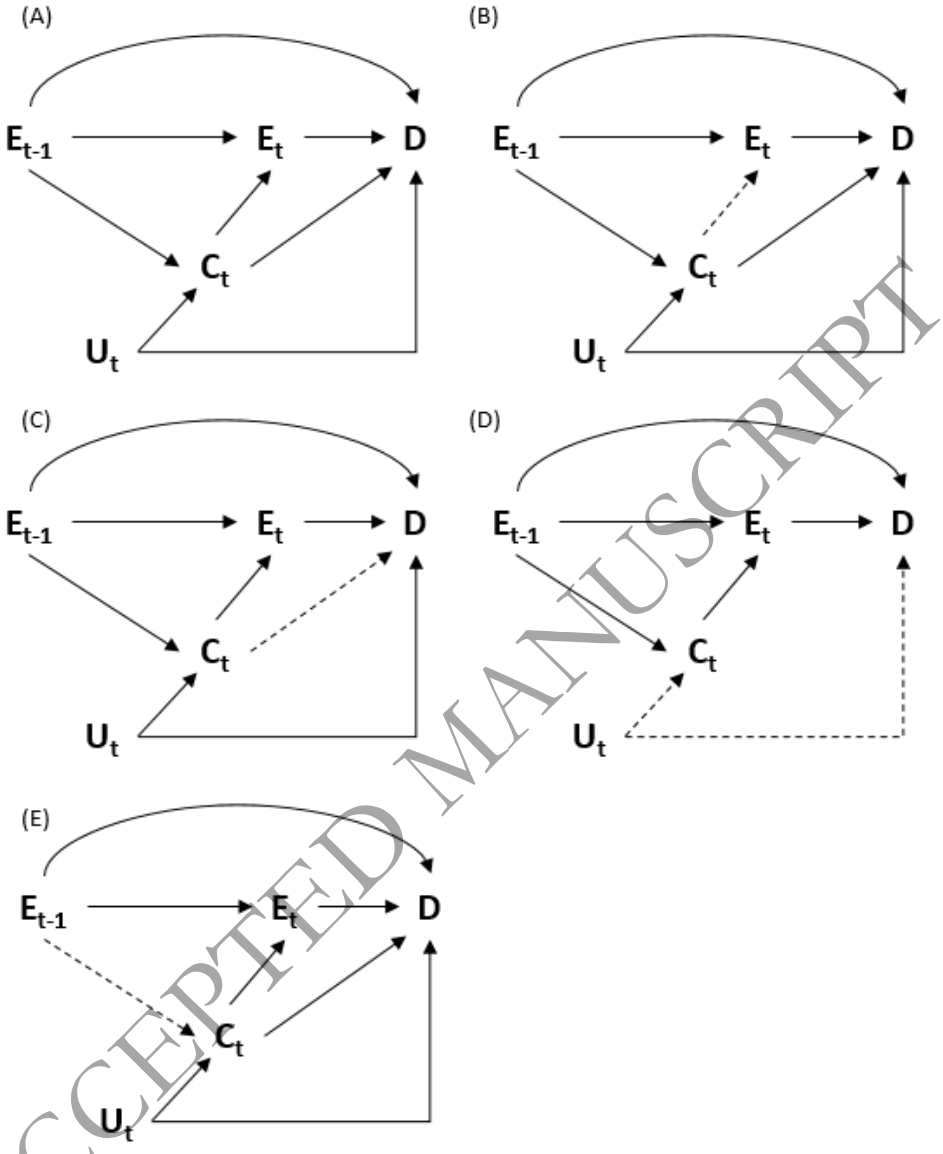
Table 3. Hazard ratios (from 8 separate analyses) for death from lymphohematopoietic cancers or female breast cancer associated with employment duration, in an occupational cohort of 13,215 employees (7549 women) at U.S. sterilization facilities, 1960-2021. Estimates are adjusted for race, plant, sex (for lymphohematopoietic cancers), calendar year, year of cohort entry, and cumulative exposure to ethylene oxide up to the previous year.

	HR for lymphohematopoietic cancer mortality	95% CI			HR for breast cancer mortality	95% CI	
Time-varying employment duration							
1-year increase in employment duration	0.98	0.96	1.00		0.97	0.95	0.99
Quartiles of cumulative employment duration				Quartiles of cumulative employment duration			
<2.9 years	Referent			<2.9 years	Referent		
[2.9, 8.9) years	0.85	0.55	1.31	[2.9, 7.6) years	0.96	0.63	1.46
[8.9, 18.3) years	0.76	0.48	1.19	[7.6, 16.0) years	0.96	0.62	1.49
≥18.3 years	0.64	0.39	1.05	≥16.0 years	0.64	0.39	1.04

Table 4. Hazard ratios for employment termination associated with quartile of ethylene oxide exposure accrued during the prior year, among employees at U.S. sterilization facilities, 1960-2021. Estimates are adjusted for race, plant, sex, calendar year, year first employed, and previous exposures in lag years 2-5.

Whole cohort (N=13,215)				Women only (N=7549)			
	HR for employment termination	95% CI			HR for employment termination	95% CI	
Ethylene oxide exposure in prior year (quartiles)				Ethylene oxide exposure in prior year (quartiles)			
< 140 ppm-days	Referent			< 126 ppm-days	Referent		
[140, 647) ppm-days	1.92	1.82	2.02	[126, 524) ppm-days	2.15	2.00	2.30
[647, 2450) ppm-days	1.41	1.33	1.50	[524, 1662) ppm-days	1.73	1.60	1.87
≥ 2450 ppm-days	1.39	1.27	1.51	≥ 1662 ppm-days	1.95	1.76	2.16

Figure 1. A simplified causal directed acyclic graph illustrating the healthy worker survivor effect. E_t refers to exposure to ethylene oxide during time interval t (*e.g.*, either average intensity during this interval or amount accrued during this interval); C_t represents employment status at time t, D is death from the cause of interest, and U_t stands for an unmeasured health status at time t. Other confounders, such as race, sex, plant, calendar year, and year of hire are omitted for clarity. Each arrow represents the possibility that the variable at the point of origin could cause a change in the variable the arrow points into. (A) The hypothesized structure of the data. In the subsequent diagrams, we use dotted arrows to draw attention to specific causal associations: (B) The dotted arrow illustrates the fact that employment affects exposure: if someone is not employed at time t, then they will not be occupationally exposed to ethylene oxide. This is known and does not need to be estimated. (C) The dotted arrow here represents the possibility that employment could directly affect risk of dying of the cause of interest. This needs to be assessed empirically in the data. If an association is found between C_t and D, then this arrow is one possible explanation for that association—*i.e.*, C_t is a shared cause of exposure and outcome, and thus a confounder. (D) Even if C_t is not a cause of D, it could still be a confounder, if it shares a cause with D. The dotted arrows here show another possible explanation for an association between C_t and D. The situations in (C) and (D) are empirically indistinguishable, but in either case (together with (B), which is known), C_t would be considered a time-varying confounder of the association between E_t and D, satisfying Criterion 1. (E) The dotted arrow here represents the possibility that prior exposure could affect employment, our Criterion 2. This needs to be assessed empirically in the data. (The criteria also depend on the directions of the associations evaluated.)



ALT TEXT FOR FIGURE 1: Five diagrams depicting causal relationships between variables E at time t-minus-1, E at time t, C at time t, U at time t, and D. In Figure 1A, there are arrows pointing from E at time t-minus-1 to 3 variables: E at time t, C at time t, and D. There are also arrows pointing from U at time t to both C at time t and D, and from C at time t to both E at time t and D. Finally there is an arrow pointing from E at time t to D. Figure 1B shows the same diagram as Figure 1A, except that the arrow from C at time t to E at time t is highlighted. Figure 1C shows the same diagram as Figure 1A, except that the arrow from C at time t to D is highlighted. Figure 1D shows the same diagram as Figure 1A, except that the arrow from C at time t to D is omitted, and both arrows from U at time t are highlighted. Figure 1E shows the same diagram as Figure 1A, except that the arrow from E at time t-minus-1 to C at time t is highlighted.