**Type of manuscript:** Original Research Article

**Manuscript title:** Evaluating hypothetical limits on metalworking fluid exposure for reducing non-Hodgkin lymphoma incidence: An application of the hazard-extended parametric g-formula

**Authors:** Kevin T. Chena,b, Sally Picciottob, Ellen A. Eisenb

a University of California, Berkeley School of Public Health, Division of Epidemiology & Biostatistics

b University of California, Berkeley School of Public Health, Division of Environmental Health Sciences

**Corresponding author:**  
Kevin T. Chen  
2121 Berkeley Way  
Room 5302, Desk 5305-1  
Berkeley, CA 94720-7360  
[kevchen@berkeley.edu](mailto:kevchen@berkeley.edu)  
(510) 387-3327

**Running head suggestion:** Limits on metalworking fluid for preventing NHL

**Conflicts of interests:** Authors declare no conflicts of interest

**Sources of financial support:** Supported by Training Grant T42OH008429, funded by the National Institute for Occupational Safety and Health (NIOSH) / Centers for Disease Control and Prevention (CDC)

**Data and computing code:** Code for reproducing analyses and reports are available on [Github](https://github.com/kvntchn/gm-nhl-ice). Data are available upon reasonable request.

**Acknowledgements:**

# Abstract

(Limit: 250 words)

**Background:** Non-Hodgkin lymphoma (NHL) incidence has increased significantly since 1960 and was recently linked with workplace exposure in the United Auto Workers-General Motors (UAW-GM) Cohort Study. Here, we investigate hypothetical interventions on workplace exposure to soluble metalworking fluids (MWF) in relation to NHL risk. Interventions on entire classes of occupational exposures rather than particular agents may result in stronger lasting protections for worker health while preventing regrettable replacements.

**Methods:** We estimated counterfactual risk of NHL from 1985 to 2004 in the UAW-GM cohort of hourly autoworkers exposed to straight, soluble, and synthetic MWF at 3 Michigan plants under hypothetical interventions. We contrasted counterfactual risk under stochastic interventions on average annual exposure to soluble MWF to that under no intervention on exposure. These stochastic interventions included (1) static interventions on soluble MWF that ignore exposure to other types and (2) dynamic interventions that seek to cap total exposure to any type of MWF by intervening on exposure to soluble MWF only. Static interventions may be more feasible as real-world policy, but dynamic interventions may result in stronger protections.

**Results:** During follow-up, 231 NHL cases occurred. Stronger hypothetical limits on average annual exposure to soluble MWF resulted in monotonically stronger reductions in NHL risk. Capping exposure to soluble MWF at 0.05 mg/m3 resulted in a risk ratio of 0.79 (95% CI: 0.62, 0.97). Capping total MWF exposure at 0.05 mg/m3 by intervening only on exposure to soluble MWF resulted in a risk ratio of 0.72 (95% CI: 0.48, 1.01).

**Conclusion:** We considered realistic interventions more compatible with a population health framework than deterministic interventions and showed that stronger limits on average annual exposure to soluble MWF may confer stronger protection against NHL risk.

# What this study adds

(Limit: 100 words)

Little is known about possible links between occupational exposures and non-Hodgkin lymphoma. Existing studies are generally single time-point case-control studies with limited exposure assessment. We leveraged the large study population size and quantitative exposure assessment of the United Auto Workers-General Motors Cohort Study to estimate counterfactual risk of non-Hodgkin lymphoma under realistic interventions using a novel causal estimator capable of adjusting for time-varying confounding affected by past exposure.

Non-Hodgkin Lymphoma (NHL) is the seventh most common cancer type in the United States and has garnered intense research attention because its incidence in the United States doubled between 1973 and 1994, before plateauing at around 19 per 100,000 persons.1,2 The strongest known risk factor of NHL is immunosuppression, both congenital and acquired.3,4 Modulation of immune function contributes to the pathogenesis of NHL directly by perturbing lymphocyte function or indirectly by reducing immunosurveillance against neoplasms. Immunodeficiency may also interact with infection with Epstein-Barr Virus, resulting in uncontrolled B-cell proliferation.5 Acquired immunodeficiency through infection with human immunodeficiency virus (HIV) can lead to genetic lesions and dysregulation while increasing vulnerability to infection by other oncogenic agents.6,7 However, since immunodeficiency and infection with HIV are rare, they cannot fully explain the historic rise or present burden of NHL.8 Since the rise in NHL incidence coincided with a period of rapid and extensive chemicalization in industry, agriculture and warfare, there has been stronger interest in the environmental and occupational epidemiology on NHL.9,10 A meta-analysis of 44 articles published between 1980 and 2014 found statistically significant associations between NHL and exposure to pesticides of several classes including carbamate, organophosporus, triazine, and organochlorine.11 While this meta-analysis presents striking evidence of the potential explanatory role of occupational exposures, it also highlights the limited breadth of occupational exposures studied.

Studies of NHL and other important occupational exposures including oils, solvents, and lubricants remain inconclusive.2 Furthermore, since NHL is relatively rare, nearly all studies of NHL incidence follow a single time-point case-control design.7,11 Large multicenter studies including EPILYMPH and InterLymph attained large enough numbers of cases to investigate particular NHL subtypes, but despite their impressive size and comprehensiveness, were limited by the lack of quantitative time-varying exposure information and the lack of adjustment for time-varying sources of bias.12,13 In the case of occupational studies, the healthy worker survivor effect (HWSE) is a key concern. The HWSE is the dynamic selection process by which healthier individuals remain at work, where they accumulate more occupational exposure, and less healthy individuals leave work.14 Standard analyses of occupational cohorts affected by the HWSE result in a downward bias on the exposure-outcome associations of interest, as less healthy workers typically accrue less exposure than healthier workers.

Metalworking fluids (MWF), an important class of occupational exposures, are complex mixtures of oils and chemicals used to lubricate and cool metal during manufacturing processes. During use, MWFs aerosolize and present a potential health hazard to exposed workers. The type of MWF applied depends in part on the particular metalworking operation. Straight MWF, the oldest class of MWF, is comprised mostly of mineral oil. Concerns over the carcinogenicity of MWF began as early as the 1970s, when the International Agency for Research on Cancer (IARC) classified mineral oils as carcinogenic, citing studies of occupational exposures among workers in oil, textile, and metal industries.15 The combination of health concerns, the high cost of oil in the 1970s, and the introduction of performance-enhancing water-soluble chemical additives led to the gradual replacement of straight MWF with water-based soluble and synthetic MWFs for many, but not all operations.16 The elimination of particular etiologic agents often leads to their replacement with other potentially hazardous chemical compounds, however.17,18 Although water-based MWF contain little to no mineral oil, they contain a richer cocktail of additives including chromates, cycloalkanes, phenols, organochlorines, nitrites, sulfonates, and triazines: classes of chemicals with known or plausible carcinogenicity.19–21 Hence, limits on entire classes of exposures grouped by functional use may result in stronger, more lasting protections for worker health than limits on particular etiologic agents because they will remain in place even as the composition and formulation of complex occupational exposures change.18 More than 20 years after the IARC classification of mineral oils as carcinogenic, the National Institute for Occupational Safety and Health (NIOSH) concluded that there exists substantial evidence linking all MWF exposure to several different cancers and now recommends an exposure limit of 0.5 mg/m3 for total particulate matter derived from any type of MWF.22,23

Here, we assess the effect of hypothetical MWF exposure limits on NHL cumulative incidence in the United Auto Workers-General Motors (UAW-GM) Occupational Cohort Study. In particular, we assess the effect of limits on soluble MWF that (1) ignore exposure to other types of MWF and (2) account for other types of MWF exposure by intervening more strongly on soluble MWF when exposure to all types of MWF is high. The latter approach is more consistent with the NIOSH REL, which was recommended for exposure to any type of MWF. However, taking into account workers’ exposure to different types of MWF may require closer monitoring and more intensive intervention. The large size of the study population and rich time-varying, quantitative MWF exposure data provide an opportunity to study this relatively rare cancer and evaluate realistic interventions on MWF exposure in a longitudinal cohort setting. A recent study of cancer incidence in this iconic occupational cohort found a monotonic association between NHL and exposure to soluble MWF.24 This association was based on a traditional Cox proportional hazards regression, a method vulnerable to bias due to confounding by time-varying factors affected by prior exposure. We therefore explore this association using a novel causal estimator capable of adjusting for time-varying confounding and censoring affected by past exposure. This estimator also yields marginal effect estimates of hypothetical interventions, which are more compatible with a population health framework than conditional measures.

# Methods

We estimated NHL risk from 1985 to 2005 under hypothetical limits on average annual exposure to soluble MWF by applying the novel hazard-extended iterative conditional expectation (ICE) parametric g-formula estimator.25 We leveraged time-varying quantitative MWF exposure data in tandem with employment records to adjust for the HWSE. First, we estimated the expected number of NHL cases per 1000 workers that we would observe if there were no competing risks. We contrasted this counterfactual risk to that when in addition, the hypothetical exposure limits of 0.5, 0.25, and 0.05 mg/m3 were enforced for soluble MWF over workers’ entire working lifetimes. These hypothetical limits are static interventions because they do not consider other covariates or exposures. However, since workers were often exposed to multiple types of MWF, both daily and cumulatively, we also evaluated the effect of limiting total MWF exposure to those hypothetical limits by intervening only on exposure to soluble MWF. These interventions, which take into consideration exposure to MWF of other types, are dynamic interventions on exposure to soluble MWF. All of the hypothetical interventions we evaluate are stochastic interventions rather than deterministic interventions because post-intervention exposures take on a range of values.

## Study population

The UAW-GM cohort includes all hourly workers at three automobile manufacturing plants in Michigan who had worked at least 3 years by 1985. Past papers provide detailed descriptions of the cohort.26,27 The present study population (N = 34,738) was restricted to the autoworkers who were at work in 1941 or not yet hired, missing no more than half of their employment history, and still alive at the start of follow-up. Autoworkers in the study population were followed for NHL incidence from January 1, 1985 until NHL diagnosis, death, 10 years after the end of employment records, or upon reaching the oldest observed age at death (108 years), whichever came earlier.

## Outcome and covariates

We identified incident cancers in the UAW-GM Cohort that occurred between 1985 and 2004 by linkage to the Michigan Cancer Registry (MCR). Workers at Plants 1 and 2, located in the greater Detroit metropolitan area, were also linked to the Detroit Regional Registry of the Surveillance, Epidemiology, and End Results (SEER) Program. Cancer types were distinguished using site and histology codes conforming to the International classification of Diseases for Oncology, 3rd edition (ICD-O-3). Non-Hodgkin lymphoma was defined by cancers with any of the following ICD-O-3 Histology codes: 9590-9597, 9670-9671, 9673, 9675, 9678-9680, 9684, 9687-9691, 9695, 9698-9702, 9705, 9708-9709, 9712, 9714-9719, 9724-9729, 9735, 9737-9738, 9811-9818, 9823, 9827, 9837. Details regarding cancer incidence follow-up are described elsewhere.24 Vital status was ascertained from company records and by linkage to Social Security Administration, National Death Index, and state mortality files.

Covariates including year of hire, sex, race, and plant location were obtained from company records. Race was missing for about 16% of the cohort, most commonly among workers hired before 1960 in Plant 2. In analyses, missing race was considered a distinct category.

## Exposure

Company industrial hygienists collected several hundred personal and area samples for total particulate matter (mg/m3) composed of MWF over many decades. Research industrial hygienists collected additional air sampling data when the cohort study was launched in the mid 1980s. These data were combined with the historical data to derive quantitative 8-hour time-weighted average exposure estimates to soluble, straight, and synthetic MWFs for each combination of job, department, and plant over time. Workers’ time-weighted average annual exposure to each MWF type was determined by combining this job-exposure matrix with employment records, which recorded time-varying job type, department, and plant. For employment records that were at least half complete, gaps in the record were interpolated by carrying forward the last known job type. The exposure assessment is described in detail elsewhere.28–30 Previous analyses of NHL applied exposure lags of 1 to 20 years; we lagged cumulative MWF exposures by 10 years to account for disease latency.31–33

## Statistical methods

We applied the hazard-extended ICE parametric g-formula with pooling over treatment history25 to estimate 20-year counterfactual risk of non-Hodgkin lymphoma. We considered six stochastic interventions on average annual exposure to soluble MWFs after the elimination of censoring/competing risks. We considered hypothetical exposure limits starting with the NIOSH Recommended Exposure Limit (REL) for total particulate mass (PM) from MWFs (0.5 mg/m3). Then, we estimated the risk at half the REL (0.25 mg/m3), and a tenth of the REL (0.05 mg/m3).22 The first three interventions were static interventions which enforced these hypothetical limits on average annual exposure to soluble MWF. The remaining three interventions were dynamic interventions that reduced average annual exposure to soluble MWF with the goal of enforcing the hypothetical limits for total exposure to MWF, regardless of type. If those limits on total average annual exposure were not possible by intervention on exposure to soluble MWF alone, average annual exposure to soluble MWF was reduced to 0. Interventions were applied at hire, before the start of follow-up, through the end of follow-up. Person-years were not intervened upon if the observed average annual exposure was below the hypothetical limit. Figure 1 illustrates the static and dynamic interventions for three person-year examples. Panel A shows the observed levels of exposure to soluble MWF and to MWF of other types. Panel B shows the post-intervention levels of exposure under a static intervention capping exposure to soluble MWF at the hypothetical limit (dotted line). Panel C shows the post-intervention levels under a dynamic intervention reducing exposure to soluble MWF with the goal of capping total MWF exposure at the hypothetical limit. Note that in person-year 3, the observed level of exposure to total MWF (and thus, to soluble MWF) is below the hypothetical limit, so the post-intervention exposure is left at the observed level.

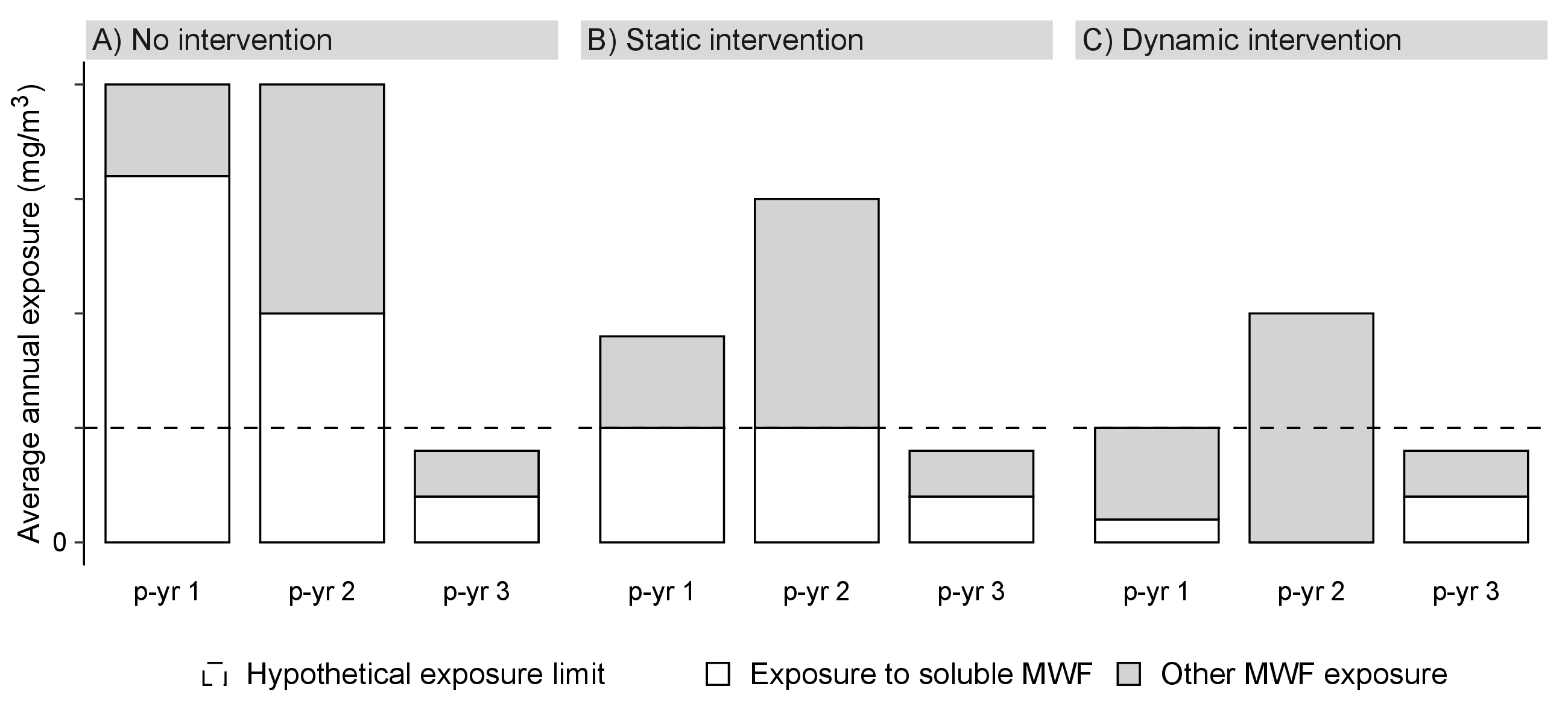


Figure 1. Observed and post-intervention levels of exposure to soluble MWF and MWF of other types in three example person-years. The white boxes represent exposure to soluble MWF, and the shaded boxes represent exposure to MWF of other types. The horizontal dotted line represents the hypothetical exposure limit.

The hazard-extended ICE parametric g-formula estimator is an ICE expression of the parametric g-formula estimator further extended to incorporate the predicted hazard of past outcomes.25 The classical ICE g-formula estimator is algebraically equivalent to the noniterative conditional expectation (NICE) g-formula estimator, which typically involves computing covariate-standardized estimates sequentially, starting at the index time, by specifying parametric models for the joint distribution of the confounders, exposures, and outcomes at each time point.34 In contrast, ICE g-formula estimators avoid the need for modeling the entire joint distribution of observed variables and only require outcome models for each time point. Counterfactual outcome estimates over the follow-up period are computed from interval-specific conditional estimates by applying the tower rule of expectation. Under the assumptions of conditional exchangeability at all time points, positivity, counterfactual consistency, and correct model specification, the hazard-extended parametric g-formula yields unbiased estimates of counterfactual risk with greater statistical efficiency than both propensity score-based estimators and the classical ICE g-formula.

Our implementation of the hazard-extended ICE parametric g-formula involved a series of model-based standardization steps using logistic regression. Post-intervention estimates of the discrete hazard of NHL given all exposures and covariates were combined iteratively from the end of follow-up to the start. In each iteration, predicted discrete hazards were standardized over post-intervention exposure and covariate histories before combining with discrete hazards from the previous iteration. This iterative process resulted in sequentially standardized estimates of NHL risk over the entire follow-up period. Averaging over the baseline distribution of covariates yielded the counterfactual risk estimate of NHL when the intervention of interest was enforced for the entire study population.

We split the 20-year follow-up period into eight time periods; the first two periods spanned four years each, and the remaining six periods spanned two years each. The first two periods are longer in length to account for the smaller number of cases in those years. Post-intervention exposure and exposure history were summarized as cumulative exposure. We modeled discrete hazards by fitting a pooled logistic regression for NHL over at-risk person-periods given cumulative exposure to straight, soluble, and synthetic MWFs, employment status, cumulative time off, year of hire, sex (male/female), race (Black/white/unknown), and plant (Plant 1/Plant 2/Plant 3). Cumulative exposure to MWFs, employment status, and cumulative time off were lagged 10 years. All continuous variables were represented as categorical variables with cut points determined by the tertiles of nonzero values among NHL cases. During the iterative combination of discrete hazards, we performed model-based standardization over baseline covariates and the complete set of time-varying covariate histories.

We estimated risk under the observed distribution of soluble MWF exposure (natural course) and under the six interventions. We contrasted the risk under intervention to that under the natural course by computing relative risks. Confidence intervals were computed using the nonparametric bootstrap with 1000 samples and centering on the estimate computed from observed data.

# Results

Table 1 presents summary statistics of exposure and covariates for the full study population and for those diagnosed with NHL between 1985 and 2004. The cohort is predominantly white (66%) and male (87%). The median year of hire among those diagnosed with NHL was 1959 whereas the median year of hire in the full study population was almost a decade later. Age at hire was approximately the same among those with NHL and the full study population. Median lagged cumulative exposure to all three MWF types was higher among NHL cases. Soluble MWFs were the most widely used MWF type, with approximately 90% of workers ever exposed. Median cumulative exposure among the exposed was 6.5 times higher for soluble than for straight MWFs. Figure 2 shows median average annual exposure to the three MWF types among exposed workers over calendar time. Exposure to MWF generally followed a downward trend over time.

Table 1: Summary of population characteristics. Statistics shown above the horizontal line are count (%). Those shown below are median (quartile 1, quartile 3).

|  | Study population | |  | NHL cases | |
| --- | --- | --- | --- | --- | --- |
| N (person-years) | 34,734 | (596,698) |  | 231 | (2,777) |
| Race |  |  |  |  |  |
| White | 22,789 | (66%) |  | 173 | (75%) |
| Black | 6,304 | (18%) |  | 21 | (9%) |
| Unknown | 5,641 | (16%) |  | 37 | (16%) |
| Sex |  |  |  |  |  |
| Male | 30,235 | (87%) |  | 206 | (89%) |
| Female | 4,499 | (13%) |  | 25 | (11%) |
| Planta |  |  |  |  |  |
| Plant 1 | 8,721 | (25%) |  | 68 | (29%) |
| Plant 2 | 14,258 | (41%) |  | 90 | (39%) |
| Plant 3 | 11,755 | (34%) |  | 73 | (32%) |
| Ever exposed to MWFsb |  |  |  |  |  |
| Straight | 19,905 | (57%) |  | 133 | (58%) |
| Soluble | 31,044 | (89%) |  | 210 | (91%) |
| Synthetic | 12,262 | (35%) |  | 72 | (31%) |
| Deceased by end of follow-up | 10,384 | (30%) |  | 33 | (14%) |
| Year of birth | 1940 | (1925, 1950) |  | 1929 | (1919, 1940) |
| Year of hire | 1967 | (1953, 1976) |  | 1959 | (1951, 1969) |
| Age at hire (years) | 23.6 | (20.0, 30.1) |  | 25.4 | (21.1, 33.6) |
| Year of leaving workc | 1979 | (1968, 1989) |  | 1977 | (1964, 1987) |
| Age at leaving work (years)c | 45.0 | (31.9, 57.7) |  | 53.4 | (36.5, 61.2) |
| Years at workc | 15.3 | (7.3, 27.1) |  | 19.2 | (8.0, 29.9) |
| Year of death | 2000 | (1993, 2008) |  | 2001 | (1994, 2005) |
| Age at death (years) | 74.7 | (65.3, 82.5) |  | 73.8 | (66.2, 82.0) |
| Cumulative time off (years)b | 1.05 | (0.30, 1.80) |  | 0.71 | (0.14, 1.40) |
| Cumulative exposure to MWFs (mg/m3-years)d | | | | | |
| Straight | 0.70 | (0.22, 2.56) |  | 0.93 | (0.29, 3.30) |
| Soluble | 4.65 | (1.85, 12.13) |  | 7.16 | (2.86, 20.91) |
| Synthetic | 0.45 | (0.16, 1.64) |  | 0.89 | (0.29, 2.11) |
| NHL: non-Hodgkin lymphoma. | | | | | |
| a Plant of longest employment duration among those who worked at multiple plants; b Lagged 10 years; c Among those who left work by December 31, 1994; d Among ever-exposed individuals, lagged 10 years. | | | | | |

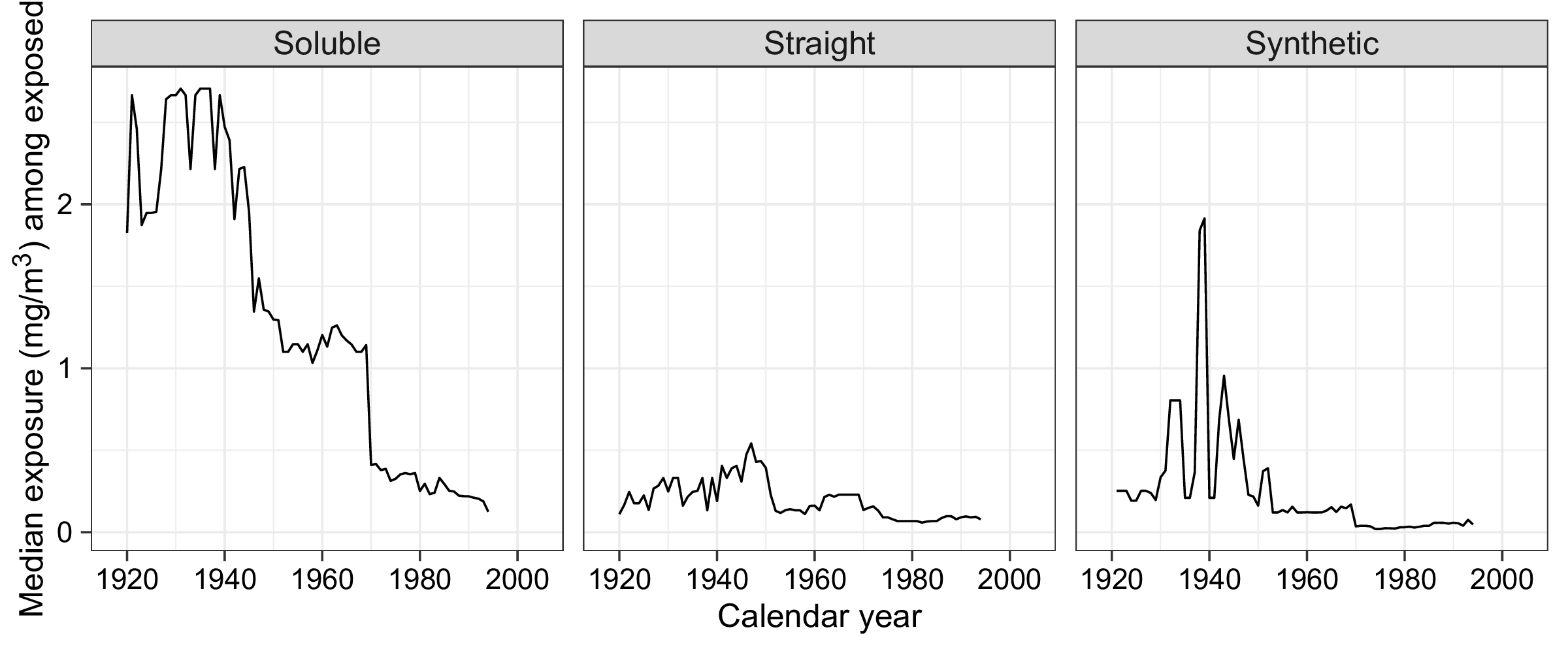


Figure 2. Median average annual exposure to soluble, straight, and synthetic metalworking fluids among exposed workers over time.

The observed risk of NHL over the 20-year follow-up period was 6.65 per 1000. Table 2 presents the hazard-extended ICE parametric g-formula estimates of the risk and risk ratios contrasting hypothetical limits on exposure to soluble MWF to no limit, after elimination of competing risks. The estimated risk under no limit on MWF exposure was 9.56 (8.15, 10.89) per 1000. Stronger limits on average annual exposure to soluble MWFs resulted in monotonically stronger reductions in the risk of NHL. Capping average annual exposure to soluble MWFs at 0.5 mg/m3, 0.25 mg/m3, and 0.05 mg/m3 resulted in a risk of 8.30 (6.52, 10.19), 8.06 (6.15, 10.15), and 7.52 (5.73, 9.51) per 1000, respectively. The risk ratios contrasting these hypothetical limits to no limit were 0.87 (0.72, 1.02), 0.84 (0.68, 1.01), and 0.79 (0.62, 0.97). Dynamic reductions in soluble exposure with the aim of capping total MWF exposure also yielded monotonically protective risk ratios: 0.84 (0.69, 0.99), 0.80 (0.64, 0.98), and 0.72 (0.48, 1.01). The dynamic interventions seeking to cap total MWF exposure at 0.5 mg/m3 and 0.25 mg/m3 attained risk reductions of similar magnitude as the static interventions enforcing (stronger) exposure limits at 0.25 and 0.05 mg/m3, respectively, but do so while intervening on a smaller proportion of person-years.

Table 2: Counterfactual risks (per 1000) and risk ratios contrasting interventions on soluble MWF to the observed course.

| Exposure limit on soluble MWF (mg/m3) | Person-years intervened (%) | Risk per 1000 | (95% CI) | RR | (95% CI) |
| --- | --- | --- | --- | --- | --- |
| None | 0.0 | 9.56 | (8.20, 10.94) | 1.00 |  |
| 0.5 | 23.8 | 8.30 | (6.44, 10.12) | 0.87 | (0.73, 1.02) |
| 0.25 | 36.2 | 8.06 | (6.00, 10.01) | 0.84 | (0.68, 1.01) |
| 0.05 | 43.9 | 7.52 | (5.57, 9.35) | 0.79 | (0.61, 0.96) |
| max(0, 0.5 – str – syn) | 28.3 | 8.00 | (6.16, 9.76) | 0.84 | (0.69, 0.99) |
| max(0, 0.25 – str – syn) | 40.0 | 7.69 | (5.78, 9.54) | 0.80 | (0.64, 0.97) |
| max(0, 0.05 – str – syn) | 52.8 | 6.89 | (4.12, 9.25) | 0.72 | (0.44, 0.97) |
| MWF: metalworking fluid; str: exposure to straight metalworking fluids; syn: exposure to synthetic metalworking fluids. | | | | | |

# Discussion

We estimated counterfactual 20-year risks of NHL from 1985 to 2005 in the UAW-GM Cohort Study under different hypothetical interventions on exposure to soluble MWFs using the hazard-extended ICE parametric g-formula. We found a monotonic exposure-dependent relationship with lower risk estimates arising from stronger limits. Contrasts in estimated NHL risk were statistically significant under the strongest static intervention, which limited soluble MWF exposure to 0.05 mg/m3 (a tenth of the current NIOSH REL), and the two dynamic interventions, which sought to limit total MWF exposure to 0.5 and 0.25 mg/m3 (the REL and half the REL, respectively) by intervening only on exposure to soluble MWF.

Under the assumptions of conditional exchangeability, positivity, consistency, and correct model specification, our estimates are unbiased for the true counterfactual risk under the hypothetical interventions. Conditional exchangeability means that for all time points, there is no confounding of the relationship between exposure/censoring and NHL status at all future time points given the observed past, including past exposure and covariates. A major threat to conditional exchangeability in longitudinal occupational studies is the HWSE. We limit potential bias due to the HWSE by conditioning on cumulative exposure, employment status, and cumulative time off history at each time point. Cumulative time off and employment status are reasonable mediators of the causal paths linking past health to future exposure and health, but adjustment for these variables may not be sufficient for eliminating bias due to the HWSE. Declines in a worker’s health may lead to reductions in work-related exposure without affecting employment status or time off work.35 We expect the absence of valid time-varying measures of worker health over the life course to result in bias toward the null.

Positivity refers to the need for adequate variation in future exposure among strata formed by observed covariate and intervention-compliant exposure histories. Even under conditional exchangeability, where exposures within these strata may be considered the result of experimental assignment, expected counterfactual outcomes under different exposures may not be estimable if there is sparsity in the observed distribution of exposures.36 We investigated static and dynamic stochastic interventions on soluble MWF exposure that intervened only when average annual exposure exceeded the hypothetical limit under consideration. Hence, our parameters of interest achieve positivity more easily than those for deterministic interventions e.g. setting all to a single level of exposure. Nonetheless, violations in positivity were still of concern due to the high dimensionality of covariates, as is common in longitudinal settings. We addressed sparsity by summarizing the 20 years of follow-up over a coarser timescale with only 8 follow-up periods and by representing covariates using fewer categories. Coarsening limits the comprehensiveness of confounding control, but improves positivity. In practice, causal inference using observational data must always balance positivity, covariate adjustment, and model specification.37 When estimating causal effects using estimators that do not require exposure modeling, as is the case with ICE g-formula methods, concerns over potential bias due to practical violations in positivity may be relaxed if correct model specification, discussed below, is attained.

The consistency assumption, also known as the no-multiple-versions-of-treatment or stable unit treatment value assumption, is that counterfactual outcomes under each possible exposure value take on a unique value.38,39 This assumption would be violated if there were multiple versions of treatment causally associated with different outcomes. This basic notion of consistency is violated in our analysis because our exposure of interest is a complex mixture of diverse components with substantial variation over time due to changes in formulation as well as the natural physical, chemical, and biological changes in the MWF over the course of its use and reuse.40 However, causal effect estimates under violations in the consistency assumption are still valid and unbiased if there is adequate adjustment for confounders of the exposure-version relationship.39 This may be thought of as conditional consistency within strata, in which there is only one version of treatment. Our analysis indexed time periods over calendar time and adjusted for age, year of hire, and plant. In this way, we limited potential for bias due to variation in MWF composition.

Correct model specification is a standard assumption in all parametric analyses. The estimator we applied offers greater statistical efficiency than the classical ICE parametric g-formula estimator because it leverages greater parametric smoothing. In causal analyses of longitudinal cohort studies, both the hazard-extended and classical ICE parametric g-formula estimators are less common than the NICE parametric g-formula.41 However, a major limitation of the NICE g-formula is the g-null paradox: the guaranteed misspecification of parametric models resulting in the false rejection of the null hypothesis when the null is true and when there is time-varying confounding affected by past exposure.42,43 As with all ICE g-formula estimators, the estimator we applied is not subject to the g-null paradox. Furthermore, simulation studies show that the variance of the hazard-extended ICE parametric g-formula is similar to that of the NICE parametric g-formula, so we expect the former to be no less conservative than the latter.25

Much of the existing epidemiologic literature linking occupational and environmental exposures to NHL report findings from case-control studies where exposures are measured crudely as binary indicators of exposure or membership in a particular occupational group.44–47 Associations between occupations and NHL risk vary considerably, but one study of working men in Kansas and Nebraska found strong associations between NHL risk and occupations involving metalworking and motor vehicles.48 Both of these occupations may entail exposure to soluble MWFs, which contain a number of additives of concern for human health and for NHL risk in particular. Organic compounds containing phosphorous, chlorine, sulfur, nitrogen, and boron are commonly added to soluble MWF to control microbial growth, improve performance under high heat/pressure, and inhibit corrosion.49 Organophosphorus compounds include organophosphate pesticides, which have been linked to cancer risk in epidemiologic and animal studies. Some were classified as possibly carcinogenic by the IARC.21 Studies of occupational exposure to chlorinated solvents and pesticides have also been linked to NHL risk.50–54 In 2014, the IARC classified trichloroethylene, tetrachloroethylene, and other chlorinated agents as Group 1 carcinogens.20 Chlorinated solvents are commonly used as degreasers in industrial settings, but their use in the plants under study here was uncommon and limited to particular operations.55 The structural characteristics shared by MWF additives and known/suspected carcinogens suggest potential similarities in their behavior in biological systems.

This study investigated the effect of hypothetical limits on MWF exposure by comparing the standardized distributions of NHL under various distributions of cumulative exposure induced by applying upper bounds to average annual exposure to soluble MWF. When average annual exposure did not exceed the hypothetical limit, it was left at the observed level of exposure. We selected these hypothetical limits based on the NIOSH REL of 0.5 mg/m3.22 In the real world, there is no enforcement of the REL, but we nonetheless observed average annual exposures below the REL in these GM plants for many years. If the REL were enforced in the real world, we would not expect reductions in exposure for these low-exposure person-years. Hence, contrasting the counterfactual scenario where all workers experienced average annual exposure at the REL to one where all workers experienced average annual exposure at some higher level would result in an overestimate of the expected real-world benefit of REL enforcement. Furthermore, since the REL is for MWF exposure of all types, we evaluated dynamic stochastic interventions that better reflect this target policy because they reduce exposure to soluble MWF with the goal of reducing MWF exposure of all types to a certain limit. The interventions we considered result in exposure distributions more similar to what we would expect in the real world; our analysis yielded results that more realistically quantify the effects of policy than those investigating deterministic interventions.56,57

Associations between several occupations and risk of NHL have been reported previously, but none evaluated the potential effect of hypothetical limits on occupational exposures.2,4,47,54 We found evidence that exposure to soluble MWF was associated with NHL incidence after adjustment for time-varying confounding affected by prior exposure using the hazard-extended ICE parametric g-formula. Reducing cumulative exposure to soluble MWF by enforcing hypothetical, but realistic, interventions on average annual exposure would reduce NHL incidence.

# Citations

1. SEER cancer statistics review 1973-1994: Trends in SEER incidence and US mortality, by race and sex. In: National Cancer Institute; 1994.

2. Ekström-Smedby K. Epidemiology and etiology of non-Hodgkin lymphoma–a review. *Acta oncologica*. 2006;45(3):258-271.

3. Filipovich A, Mathur A, Kamat D, Shapiro R. Primary immunodeficiencies: Genetic risk factors for lymphoma. *Cancer research*. 1992;52(19\_Supplement):5465s-5467s.

4. Chiu BCH, Hou N. Epidemiology and etiology of non-Hodgkin lymphoma. In: Evens AM, Blum KA, eds. Vol 165. Springer; 2015.

5. Oertel SH, Riess H, eds. *Immunosurveillance, Immunodeficiency and Lymphoproliferations: Lymphoproliferative Disordersin Congenital and Acquired Immunodeficiencies*. Springer; 2002:1-8.

6. Grulich AE, Wan X, Law MG, Coates M, Kaldor JM. Risk of cancer in people with AIDS. *Aids*. 1999;13(7):839-843.

7. Alexander DD, Mink PJ, Adami H-O, et al. The non-Hodgkin lymphomas: A review of the epidemiologic literature. *International Journal of Cancer*. 2007;120(S12):1-39.

8. Shiels MS, Engels EA, Linet MS, et al. The epidemic of non-Hodgkin lymphoma in the United States: Disentangling the effect of HIV, 1992–2009. *Cancer Epidemiology and Prevention Biomarkers*. 2013;22(6):1069-1078.

9. Nelson NJ. Studies examine whether persistent organic agents may be responsible for rise in lymphoma rates. *Journal of the National Cancer Institute*. 2005;97(20):1490-1491.

10. Romero AM. *Economic Poisoning: Industrial Waste and the Chemicalization of American Agriculture*. University of California Press; 2021. doi:[doi:10.1525/9780520381575](https://doi.org/doi:10.1525/9780520381575).

11. Schinasi L, Leon ME. Non-Hodgkin lymphoma and occupational exposure to agricultural pesticide chemical groups and active ingredients: A systematic review and meta-analysis. *International Journal of Environmental Research and Public Health*. 2014;11(4):4449-4527. doi:[10.3390/ijerph110404449](https://doi.org/10.3390/ijerph110404449).

12. Besson H, Brennan P, Becker N, et al. Tobacco smoking, alcohol drinking and non-Hodgkin’s lymphoma: A European multicenter case-control study (epilymph). *International journal of cancer*. 2006;119(4):901-908.

13. Morton LM, Sampson JN, Cerhan JR, et al. Rationale and design of the international lymphoma epidemiology consortium (InterLymph) non-Hodgkin lymphoma subtypes project. *J Natl Cancer Inst Monogr*. 2014;2014(48):1-14. doi:[10.1093/jncimonographs/lgu005](https://doi.org/10.1093/jncimonographs/lgu005).

14. Arrighi HM, Hertz-Picciotto I. The evolving concept of the healthy worker survivor effect. *Epidemiology*. 1994;5(2):189-196. <http://www.jstor.org/stable/3702361>.

15. IARC. *IARC Monographs on the Evaluation of Carcinogenic Risk of the Chemical to Man: Certain Polycyclic Aromatic Hydrocarbons and Heterocyclic Compounds*. Vol 3. World Health Organization International Agency for Research on Cancer; 1973.

16. Childers J. The chemistry of metalworking fluids. In: *Metalworking Fluids*. CRC Press; 2006.

17. Kassotis CD, Vandenberg LN, Demeneix BA, Porta M, Slama R, Trasande L. Endocrine-disrupting chemicals: Economic, regulatory, and policy implications. *The lancet Diabetes & endocrinology*. 2020;8(8):719-730.

18. Maertens A, Golden E, Hartung T. Avoiding regrettable substitutions: Green toxicology for sustainable chemistry. *ACS Sustainable Chemistry & Engineering*. 2021;9(23):7749-7758.

19. IARC. *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans: Overall Evaluations of Carcinogenicity: An Updating of IARC Monographs*. Vol 1-42. World Health Organization International Agency for Research on Cancer; 1987:106-116.

20. IARC. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Trichloroethylene, Tetrachloroethylene, and Some Other Chlorinated Agents*. Vol 106. World Health Organization International Agency for Research on Cancer; 2014:106-116.

21. IARC. *IARC Monographs on the Evaluation of Carcinogenic Risk of the Chemical to Humans: Some Organophosphate Insecticides and Herbicides*. Vol 112. World Health Organization International Agency for Research on Cancer; 2017.

22. Rosenstock L, ed. *What You Need to Know about Occupational Exposure to Metalworking Fluids*. Department of Health; Human Services (NIOSH); 1998.

23. Mirer F. Updated epidemiology of workers exposed to metalworking fluids provides sufficient evidence for carcinogenicity. *Applied occupational and environmental hygiene*. 2003;18(11):902-912.

24. Colbeth HL, Chen KT, Picciotto S, Costello S, Eisen EA. Metalworking fluids and cancer incidence in the United Autoworkers-General Motors Cohort. *American Journal of Epidemiology*. 2022, in press.

25. Wen L, Young JG, Robins JM, Hernán MA. Parametric g-formula implementations for causal survival analyses. *Biometrics*. 2020.

26. Eisen EA, Tolbert PE, Monson RR, Smith TJ. Mortality studies of machining fluid exposure in the automobile industry I: A standardized mortality ratio analysis. *American journal of industrial medicine*. 1992;22(6):809-824.

27. Eisen EA, Bardin J, Gore R, Woskie SR, Hallock MF, Monson RR. Exposure-response models based on extended follow-up of a cohort mortality study in the automobile industry. *Scandinavian journal of work, environment & health*. 2001;27(4):240-249.

28. Hallock MF, Smith TJ, Woskie SR, Hammond SK. Estimation of historical exposures to machining fluids in the automotive industry. *American Journal of Industrial Medicine*. 1994;26(5):621-634. doi:[10.1002/ajim.4700260505](https://doi.org/10.1002/ajim.4700260505).

29. 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. *American Industrial Hygiene Association Journal*. 1994;55(1):20-29.

30. Woskie SR, Virji MA, Hallock M, Smith TJ, Hammond SK. Summary of the findings from the exposure assessments for metalworking fluid mortality and morbidity studies. *Applied occupational and environmental hygiene*. 2003;18(11):855-864.

31. Smith MT, Jones RM, Smith AH. Benzene exposure and risk of non-Hodgkin lymphoma. *Cancer epidemiology, biomarkers & prevention*. 2007;16(3):385-391.

32. Karipidis KK, Benke G, Sim MR, et al. Occupational exposure to ionizing and non-ionizing radiation and risk of non-Hodgkin lymphoma. *International archives of occupational and environmental health*. 2007;80(8):663-670.

33. Zhang L, Rana I, Shaffer RM, Taioli E, Sheppard L. Exposure to glyphosate-based herbicides and risk for non-Hodgkin lymphoma: A meta-analysis and supporting evidence. *Mutation Research/Reviews in Mutation Research*. 2019;781:186-206.

34. Taubman SL, Robins JM, Mittleman MA, Hernán MA. Intervening on risk factors for coronary heart disease: An application of the parametric g-formula. *International journal of epidemiology*. 2009;38(6):1599-1611.

35. 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. *Occupational and environmental medicine*. 2017;74(4):294-300.

36. Maldonado G, Greenland S. Estimating causal effects. *International journal of epidemiology*. 2002;31(2):422-429.

37. Petersen ML, Porter P, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. *Statistical Methods in Medical Research*. 2012;21(1):31-54. doi:[10.1177/0962280210386207](https://doi.org/10.1177/0962280210386207).

38. Cole SR, Frangakis CE. The consistency statement in causal inference: A definition or an assumption? *Epidemiology*. 2009;20(1):3-5.

39. VanderWeele TJ, Shpitser I. On the definition of a confounder. *Annals of statistics*. 2013;41(1):196.

40. Howell JK, Lucke WE, White EM. Metalworking fluids. In: Byers JP, ed. CRC Press; 2006.

41. Keil AP, Edwards JK, Richardson DB, Naimi AI, Cole SR. The parametric g-formula for time-to-event data: Intuition and a worked example. *Epidemiology*. 2014;25(6). <https://journals.lww.com/epidem/Fulltext/2014/11000/The_Parametric_g_Formula_for_Time_to_event_Data_.16.aspx>.

42. Naimi AI, Tchetgen Tchetgen EJ. Invited commentary: Estimating population impact in the presence of competing events. *American journal of epidemiology*. 2015;181(8):571-574.

43. McGrath S, Young JG, Hernán MA. Revisiting the g-null paradox. *Epidemiology*. 2022;33(1):114-120.

44. Cano MI, Pollán M. Non-Hodgkin’s lymphomas and occupation in sweden. *International archives of occupational and environmental health*. 2001;74(6):443-449.

45. Costantini AS, Miligi L, Kriebel D, et al. A multicenter case-control study in italy on hematolymphopoietic neoplasms and occupation. *Epidemiology*. 2001:78-87.

46. Karunanayake CP, McDuffie HH, Dosman JA, Spinelli JJ, Pahwa P. Occupational exposures and non-Hodgkin’s lymphoma: Canadian case-control study. *Environmental Health*. 2008;7(1):1-9.

47. ‘t Mannetje A, De Roos AJ, Boffetta P, et al. Occupation and risk of non-Hodgkin lymphoma and its subtypes: A pooled analysis from the InterLymph consortium. *Environmental health perspectives*. 2016;124(4):396-405.

48. Zheng T, Blair A, Zhang Y, Weisenburger DD, Zahm SH. Occupation and risk of non-Hodgkin’s lymphoma and chronic lymphocytic leukemia. *Journal of occupational and environmental medicine*. 2002;44(5):469-474.

49. Evans R, Hooijman J, van der Veer J. High-speed machining. In: Gupta K, Davim P, eds. Academic Press; 2020.

50. Cocco P, Brennan P, Ibba A, et al. Plasma polychlorobiphenyl and organochlorine pesticide level and risk of major lymphoma subtypes. *Occupational and Environmental Medicine*. 2008;65(2):132-140.

51. Purdue MP, Bakke B, Stewart P, et al. A case-control study of occupational exposure to trichloroethylene and non-Hodgkin lymphoma. *Environmental health perspectives*. 2011;119(2):232-238.

52. Cocco P, Vermeulen R, Flore V, et al. Occupational exposure to trichloroethylene and risk of non-Hodgkin lymphoma and its major subtypes: A pooled IinterLlymph analysis. *Occupational and environmental medicine*. 2013;70(11):795-802.

53. Vlaanderen J, Straif K, Pukkala E, et al. Occupational exposure to trichloroethylene and perchloroethylene and the risk of lymphoma, liver, and kidney cancer in four nordic countries. *Occupational and environmental medicine*. 2013;70(6):393-401.

54. Callahan CL, Stewart PA, Friesen MC, et al. Case-control investigation of occupational exposure to chlorinated solvents and non-Hodgkin’s lymphoma. *Occupational and environmental medicine*. 2018;75(6):415-420.

55. Shrestha D, Liu S, Hammond SK, et al. Risk of renal cell carcinoma following exposure to metalworking fluids among autoworkers. *Occupational and environmental medicine*. 2016;73(10):656-662.

56. Bembom O, van der Laan MJ. A practical illustration of the importance of realistic individualized treatment rules in causal inference. *Electronic journal of statistics*. 2007;1:574.

57. Muñoz ID, van der Laan M. Population intervention causal effects based on stochastic interventions. *Biometrics*. 2012;68(2):541-549.