**Type of manuscript:** Original Research Article

**Manuscript title:** Hypothetical limits on workplace exposure for reducing non-Hodgkin lymphoma: An application of the hazard-extended iterative conditional expectation parametric g-formula

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

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

c University of California, Berkeley Department of Statistics

**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:** This work was supported by Training Grant T42OH008429 and Research Project Grant R01OH011092, both 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.

**Keywords:** Metalworking fluid, occupational exposure, non-Hodgkin lymphoma, cancer, causal inference, g-formula, healthy-worker effect, United States

# Abstract

(Limit: 250 words)

**Background:** Non-Hodgkin lymphoma (NHL) incidence has increased significantly since 1960 and was recently linked to workplace exposure to soluble metalworking fluids (MWF) in a standard survival analysis of the United Auto Workers-General Motors (UAW-GM) cohort. Here, we investigate the causal effects of hypothetical limits on soluble MWF exposure in relation to NHL risk using an iterative conditional expectation (ICE) parametric g-formula estimator.

**Methods:** We estimated counterfactual risk of NHL between 1985 and 2005 in the UAW-GM cohort of autoworkers at three Michigan plants (n = 34,734) under hypothetical limits on average annual exposure to soluble MWF. We applied the hazard-extended ICE parametric g-formula, an estimator capable of adjusting for time-varying confounding affected by past exposure.

**Results:** During follow-up, 231 NHL cases occurred. Stronger hypothetical limits on soluble MWF exposure resulted in monotonic reductions in NHL risk estimates. Capping average annual exposure at 0.5 mg/m3, the NIOSH recommended exposure limit, would have prevented 44 (95% CI: -6, 91) cases. Capping at 0.25 mg/m3 would have prevented 52 (95% CI: -5, 106) cases. Finally, capping at 0.05 mg/m3 would have prevented 71 (95% CI: 12, 129) cases.

**Conclusion:** Stronger limits on soluble MWF exposure would have prevented more cases of NHL in the UAW-GM cohort. Our application demonstrates the utility of an ICE g-formula estimator for estimating the effect of realistic exposure interventions on a survival outcome. Greater availability of worked examples and software would facilitate wider application of ICE g-formula estimators.

Non-Hodgkin Lymphoma (NHL) is the seventh most common cancer type in the United States and has garnered substantial research attention because its incidence in the United States doubled between 1973 and 1994, before plateauing at around 19 per 100,000 persons per year.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

Pesticide exposure in agricultural occupational settings was a common target of NHL research in recent decades. 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 Occupations associated with NHL are not limited to the agricultural sector, however. Metal processors, health workers, salespeople, machinists, and electricians have all been identified as occupational groups with higher risk of NLH.2,12,13 Studies contrasting occupational groups to the general population are vulnerable to exposure misclassification of the true causal agent. When different causal agents contribute to overall NHL risk by their effect on particular subtypes of NHL, their etiologic role may be obscured in analyses not stratified by subtype. Large multicenter studies including EPILYMPH and INTERLYMPH seek to disentangle effect heterogeneity across NHL subtypes, but are limited by the lack of quantitative time-varying exposure information.14–16

Measures of time-varying exposure are essential to statistical adjustment for forms of time-varying confounding affected by past exposure such as the healthy worker survivor effect (HWSE). 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.17 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 oils, which were classified as carcinogenic by the International Agency for Research on Cancer (IARC) in 1973, citing studies of occupational exposures among workers in oil, textile, and metal industries.18 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.19

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.20–22 The elimination of particular etiologic agents often leads to their replacement with other potentially hazardous chemical compounds.23,24 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.24 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. However, the NIOSH recommended exposure limit of 0.5 mg/m3 for total particulate matter derived from any type of MWF was intended to reduce respiratory disorders and not cancer.25,26

Here, we assess the effect of hypothetical MWF exposure limits on NHL cumulative incidence (which we will refer to as risk) from 1985 to 2005 in the United Auto Workers-General Motors (UAW-GM) Occupational Cohort Study. 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.27 This association was based on standard Cox proportional hazards regression, a method vulnerable to bias due to confounding by time-varying factors affected by past exposure. We therefore explore this association using an estimator capable of adjusting for time-varying confounding and censoring affected by past exposure: the hazard-extended iterative conditional expectation (ICE) parametric g-formula. Like other causal estimators in statistics, the hazard-extended ICE g-formula yields marginal effect estimates of hypothetical interventions, which are more compatible with a population health framework than conditional measures.

The g-formula is a well-known result in causal inference for the identification of causal effects in the presence of time-varying confounding affected by past exposure.28 Standard representations of the g-formula include (1) a non-iterated expectation over the joint density of covariates, (2) the ICE over time, and (3) an inverse probability weighted expectation. The parametric g-formula is a plug in estimator for the g-formula under its first, non-iterative, representation. It involves the parametric modeling of the full joint distribution of the outcome, exposure, and covariates under study for each time point.29,30 Counterfactual quantities under hypothetical interventions of interest are computed from Monte Carlo samples from distributions implied by the fitted parametric models. In longitudinal settings, this approach often requires specifying and fitting large number of models in order to satisfy the exchangeability assumptions necessary for causal identification.

When researchers are not wholly comfortable with the full set of parametric assumptions required by the parametric g-formula, the ICE g-formula may be preferred. Parametric estimators using the ICE representation of the g-formula require modeling only the conditional outcome distributions, so they require fewer parametric assumptions than non-iterative parametric g-formula estimators. 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.

# Methods

We estimated NHL risk from 1985 to 2005 under hypothetical limits on average annual exposure to soluble MWF by applying the hazard-extended iterative conditional expectation (ICE) parametric g-formula estimator.31 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 that we would observe if there were no censoring by competing risks. Then, we contrasted this counterfactual risk to that when in addition, the hypothetical exposure limits of (1) the NIOSH Recommended Exposure Limit (REL) for total particulate mass (PM) from MWFs (0.5 mg/m3), (2) half the REL (0.25 mg/m3), and (3) a tenth of the REL (0.05 mg/m3) were enforced for soluble MWF over workers’ entire working lifetimes.25 Under these hypothetical interventions, exposure was set at the hypothetical limit if observed exposure was greater than that limit. Otherwise, exposure not intervened upon. This kind of intervention is known as a threshold longitudinal modified treatment policy.29,30,32,33 Longitudinal modified treatment policies are intervention strategies that depend on the natural value of exposure at a particular time ie the value of exposure that would have been observed at that time if the intervention were discontinued immediately prior.32,34

## 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.35,36 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, January 1, 2005 (10 years after the end of employment record availability), or age 108 years (the oldest observed age at death), whichever came earlier.

## Outcome and covariates

We identified incident cancers in the UAW-GM Cohort that occurred between 1985 and 2005 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. Subtypes of NHL were also identified for sensitivity analyses. Details regarding cancer incidence follow-up are described elsewhere.27 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. All covariates were coded as categorical variables. Cut-points for continuous covariates were determined according to the quantiles among cases.

## Exposure

Quantitative measurement of time-varying MWF exposure is a distinguishing strength of the UAW-GM cohort relative to other studies of occupational exposure. Exposure assessment was based on direct sampling as well as company records. 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 and company records 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 through 1994. 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.37–39 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 and therefore ended follow-up on January 1, 2005.40–42 In analyses, MWF exposure history was summarized as the cumulative sum of annual exposure intensities and coded as categorical variables with cut-points determined according the quantiles of cumulative exposure among cases.

## Statistical methods

The hazard-extended ICE parametric g-formula may be thought of as a series of model-based standardization steps, which we implemented using logistic regression. 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 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 results in sequentially standardized estimates of NHL risk over the entire follow-up period. Averaging over the baseline distribution of covariates yields the counterfactual risk estimate of NHL when the intervention of interest was enforced for the entire study population.

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. An overview of the general steps of the estimation procedure are presented below.

1. For the hypothetical exposure limit of interest, compute the cumulative exposure each worker would have accrued by the end of each follow-up period since hire.
2. Fit a pooled logistic regression to the observed data for NHL on covariates and cumulative exposure over all at-risk person-periods, excluding those ending with a censoring event.
3. Replace the cumulative exposure vector in the observed data with the post-intervention cumulative exposure vector. Using the model fitted in the previous step, compute estimates of the post-intervention discrete hazard for each at-risk person-period including those that end with a censoring event.
4. Set where is the number of follow-up periods:
   1. Among those who were event-free and uncensored through the th period, fit a logistic regression on the predicted post-intervention discrete hazard for the period spanning the through th periods given observed covariate and exposure history up through (and including) the th period.
   2. Replace the cumulative exposure vector in the observed data with the post-intervention cumulative exposure vector. Using the model fitted in the previous step, obtain predicted values for all those who were at-risk in the th period.
   3. Compute the predicted post-intervention discrete hazard for the period spanning the through th periods by multiplying the predicted values from the previous step by 1 minus the discrete hazard estimate used to fit the model in step 3.a and adding that same discrete hazard estimate to the product.
   4. If , set and return to step 3.a.
5. Compute the counterfactual risk by averaging the predicted post-intervention discrete hazards for the entire follow-up period for all units.

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 Monte Carlo samples from the population at risk at start of follow-up and centering on the estimate computed from observed data. All the necessary script used to reproduce the analyses are available on [GitHub](https://github.com/kvntchn/gm-nhl-ice).

# Results

Table 1 presents summary statistics of exposure and covariates for the full study population and for those diagnosed with NHL between 1985 and 2005. 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 | 1.07 | (0.33, 1.87) |  | 0.68 | (0.17, 1.43) |
| 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.; | | | | | |

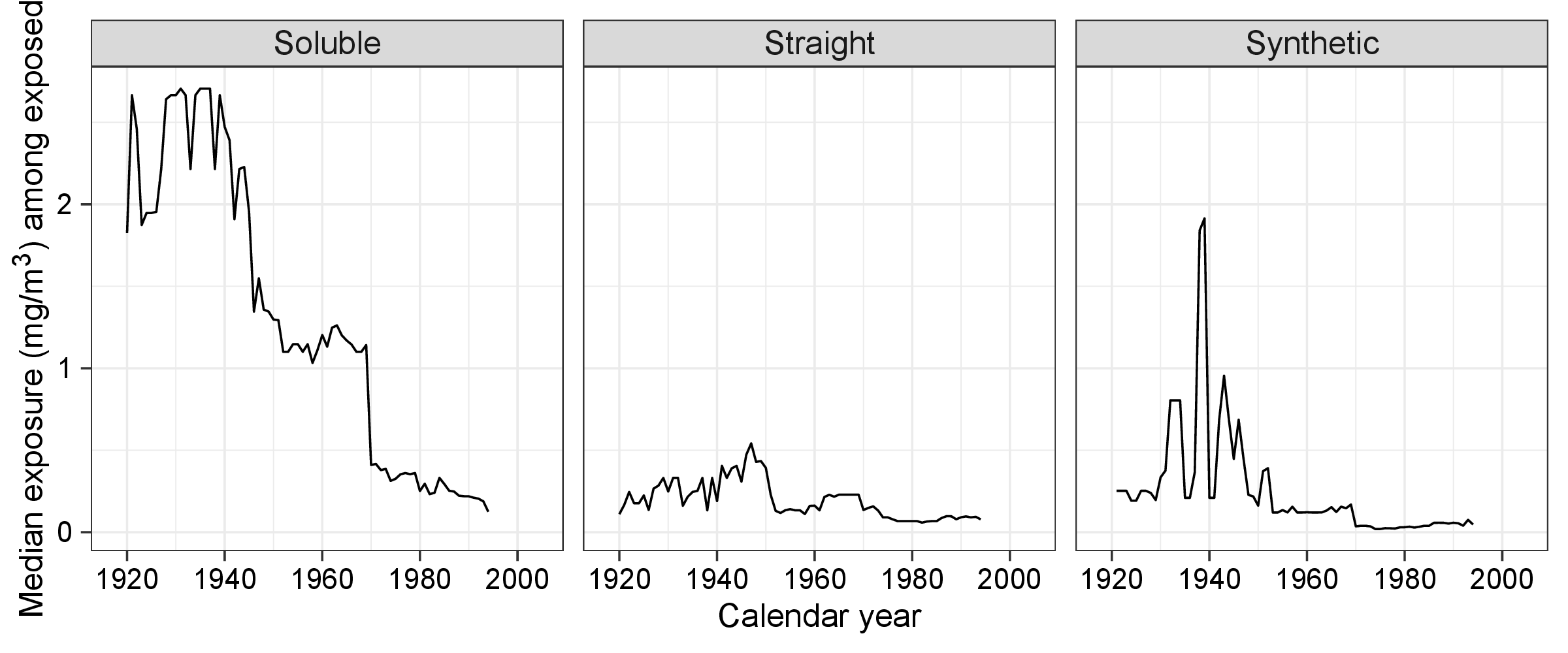


Figure 2. Median average annual exposure to soluble, straight, and synthetic metalworking fluids among exposed workers over time.

The observed risk of NHL over the 20-year follow-up period was 231 per 34,734 (6.65 per 1000). Table 2 presents the hazard-extended ICE parametric g-formula estimates of the counterfactual risk, risk difference, and risk ratios contrasting hypothetical limits on exposure to soluble MWF to no limit, after elimination of competing risks. Under an intervention eliminating competing risks, the estimated risk under no limit on MWF exposure was 332 (285, 380) per 34,734. 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 44 (-6, 91), 52 (-5, 106), and 71 (12, 129) fewer NHL cases, respectively. These correspond to risk ratios 0.87 (0.73, 1.02), 0.84 (0.68, 1.01), and 0.79 (0.61, 0.97).

Table 2: Counterfactual risk, risk difference, and risk ratio estimates contrasting interventions on soluble MWF to the observed course.

| Exposure limit (mg/m3) | Person-years intervened (%) | Risk | (95% CI) | RD | (RD 95% CI) | RR | (95% CI) |
| --- | --- | --- | --- | --- | --- | --- | --- |
| None | 0.0 | 332 | (285, 380) |  |  |  |  |
| 0.5 | 23.8 | 288 | (224, 351) | 44 | (-6, 91) | 0.87 | (0.73, 1.02) |
| 0.25 | 36.2 | 280 | (209, 348) | 52 | (-5, 106) | 0.84 | (0.68, 1.01) |
| 0.05 | 43.9 | 261 | (193, 325) | 71 | (12, 129) | 0.79 | (0.61, 0.96) |
| Note. Counterfactual risk and risk difference estimates were expressed per 34,734. MWF: metalworking fluid | | | | | | | |

# 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.

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 both future exposure/censoring and NHL status given the observed past, including past exposure and covariates.32,34 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.43 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 excessive sparsity in the observed distribution of exposures.44 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 static 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.45

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.46,47 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.48 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.47 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 and semi-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.49 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.50,51 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.31

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.52–55 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.56 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.57 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.22 Studies of occupational exposure to chlorinated solvents and pesticides have also been linked to NHL risk.58–62 In 2014, the IARC classified trichloroethylene, tetrachloroethylene, and other chlorinated agents as Group 1 carcinogens.21 Chlorinated solvents are commonly used as degreasers in industrial settings, but their use in the plants under study here was rare.63 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. We selected these hypothetical limits based on the NIOSH REL of 0.5 mg/m3.25 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.

Associations between several occupations and risk of NHL have been reported previously, but none evaluated the potential effect of hypothetical limits on occupational exposures.2,4,55,62 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. Institute NC. SEER cancer statistics review 1973-1994: Trends in SEER incidence and US mortality, by race and sex. 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*. Vol 165. (Evens AM, Blum KA, eds.). 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. Fritschi L, Benke G, Hughes AM, et al. Risk of non-Hodgkin lymphoma associated with occupational exposure to solvents, metals, organic dusts and PCBs (australia). *Cancer Causes & Control*. 2005;16(5):599-607.

13. Mester B, Nieters A, Deeg E, Elsner G, Becker N, Seidler A. Occupation and malignant lymphoma: A population based case control study in germany. *Occupational and environmental medicine*. 2006;63(1):17-26.

14. 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.

15. ’t Mannetje A, Dryson E, Walls C, et al. High risk occupations for non-Hodgkin’s lymphoma in new zealand: Case–control study. *Occupational and environmental medicine*. 2008;65(5):354-363.

16. 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).

17. 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>.

18. 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.

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

20. 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.

21. 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.

22. 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.

23. 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.

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

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

26. 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.

27. 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.

28. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*. 1986;7(9):1393-1512. doi:<https://doi.org/10.1016/0270-0255(86)90088-6>.

29. 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.

30. Young JG, Hernán MA, Robins JM. Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. *Epidemiologic methods*. 2014;3(1):1-19.

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

32. Richardson TS, Robins JM. Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series Working Paper*. 2013;128(30):2013.

33. Haneuse S, Rotnitzky A. Estimation of the effect of interventions that modify the received treatment. *Statistics in medicine*. 2013;32(30):5260-5277.

34. Dı́az I, Williams N, Hoffman KL, Schenck EJ. Nonparametric causal effects based on longitudinal modified treatment policies. *Journal of the American Statistical Association*. 2021:1-16.

35. 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.

36. 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.

37. 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).

38. 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.

39. 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.

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

41. 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.

42. 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.

43. 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.

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

45. 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).

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

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

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

49. 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>.

50. 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.

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

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

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

54. 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.

55. ‘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.

56. 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.

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

58. 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.

59. 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.

60. 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.

61. 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.

62. 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.

63. 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.