# 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 covariates1,2. 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 work3. 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 exposures4. 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 specification5. 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 value6,7. 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 reuse8. 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 relationship7. 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-formula9. 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 exposure10,11. 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 latter12.

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 group13–16. 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 vehicles17. 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 corrosion18. 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 IARC19. Studies of occupational exposure to chlorinated solvents and pesticides have also been linked to NHL risk20–24. In 2014, the IARC classified trichloroethylene, tetrachloroethylene, and other chlorinated agents as Group 1 carcinogens25. 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 operations26. 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/m327. 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 exposures16,24,28,29. 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.

1. 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.

2. 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.

3. 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.

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

5. 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)

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

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

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

9. 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.](https://journals.lww.com/epidem/Fulltext/2014/11000/The_Parametric_g_Formula_for_Time_to_event_Data_.16.aspx)

10. 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.

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

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

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

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

15. 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.

16. ‘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.

17. 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.

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

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. 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.

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

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

29. Chiu BCH, Hou N. *Epidemiology and Etiology of Non-Hodgkin Lymphoma*. Vol 165. (Evens AM, Blum KA, eds.). Springer; 2015.