# Methods

## Study population

The UAW-GM Cohort included 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 (Eisen et al. 1992, 2001). The present study population (N = 34,748) was restricted to the autoworkers who were still alive by 1985 and missing no more than half of their employment history. Autoworkers in the study population were followed until December 31, 2004.

## 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, which began collecting cancer case data in 1973. 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. See Colbeth et al. (2022, in press) for details regarding cancer incidence follow-up. Vital status was ascertained by company records and linkage to the Social Security Administration, National Death Index, death certificates, and state mortality files.

Covariates including year of hire, sex, race, and plant location were obtained from company records. Race data were missing for about 16% of the workers. In analyses, missing race was considered a distinct category of race.

## Exposure

For each combination of job type, department, and plant industrial hygienists collected several hundred personal and area samples for particulate matter (mg/m3) over the course of three decades in order to derive quantitative 8-hour time-weighted average exposures to straight, soluble, and synthetic MWFs. Workers’ time-weighted annual exposure to each of the MWFs was determined according to employment records, which recorded time-varying job, department,and plant. For employment records that were no less than half complete, gaps in record were interpolated by carrying forward the last known job. Cumulative exposures to straight, soluble, and synthetic MWFs (mg/m3-years) were calculated by taking cumulative sums of annual exposures. In analyses, MWF exposure was lagged 10 years. Exposure assessment is described in detail elsewhere (Hallock et al. 1994; Woskie et al. 1994, 2003).

## Statistical methods

We applied the hazard-extended parametric g-formula with pooling over treatment history (Wen et al. 2020) to estimate 20-year counterfactual cumulative incidence of non-Hodgkin lymphoma under five hypothetical interventions on MWF exposure of a single type (straight, soluble, or synthetic) with the elimination of censoring/competing risks. The first, second, and third interventions limited average annual exposure of a particular MWF type to 0.5, 0.25, and 0.05 mg/m3, respectively. The fourth and fifth, interventions limited average annual exposure of all MWF types to 0.5 or 0.25 mg/m3 by intervening on a single MWF type, if possible. If not, average annual exposure of the particular MWF type was reduced to 0. The NIOSH relative exposure limit for total particulate mass from MWFs is 0.5 mg/m3 (Rosenstock 1998).

The hazard-extended parametric g-formula estimator is an iterated conditional expectation (ICE) expression of the parametric g-formula estimator extended to incorporate the predicted hazard of past outcomes. This differs from the classical iterated conditional expectation estimator where past outcomes are incorporated as indicators (Wen et al. 2020). Under the assumptions of correct model specification, conditional exchangeability at all time points, positivity, and counterfactual consistency, the hazard-extended parametric g-formula yields unbiased estimates of counterfactual cumulative incidences with greater statistical efficiency than both propensity-score based estimators and the classical ICE g-formula.

Our implementation is thoroughly described in the Appendix. In brief, we conducted 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 distributions of the exposure and covariates before combining with discrete hazards from the previous iteration. The result of this iterative process are sequentially-standardized estimates of the risk of NHL over the entire follow-up period. Taking the average of these estimates yields the estimated counterfactual risk of NHL when all individuals in the study population are subject to the intervention of interest.

We split the 20-year follow-up period into eight periods where the first two periods spanned four years each, and the remaining six periods spanned two years each. 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 metalworking fluids, employment status, year of hire, cumulative time off, sex (male/female), race (Black/white/unknown), and plant (Plant 1/Plant 2/Plant 3). 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 time-varying covariate histories. Confidence intervals were computed using the nonparametric bootstrap over 1000 samples with centering on the estimate computed from observed data.

Colbeth, H. L., Chen, K. T., Picciotto, S., Costello, S., and Eisen, E. A. (2022, in press), “Metalworking fluids and cancer incidence in the united autoworkers-general motors cohort,” *American Journal of Epidemiology*.

Eisen, E. A., Bardin, J., Gore, R., Woskie, S. R., Hallock, M. F., and Monson, R. R. (2001), “Exposure-response models based on extended follow-up of a cohort mortality study in the automobile industry,” *Scandinavian journal of work, environment & health*, 27, 240–249.

Eisen, E. A., Tolbert, P. E., Monson, R. R., and Smith, T. J. (1992), “Mortality studies of machining fluid exposure in the automobile industry I: A standardized mortality ratio analysis,” *American journal of industrial medicine*, Wiley Online Library, 22, 809–824.

Hallock, M. F., Smith, T. J., Woskie, S. R., and Hammond, S. K. (1994), “Estimation of historical exposures to machining fluids in the automotive industry,” *American Journal of Industrial Medicine*, Wiley Subscription Services, Inc., A Wiley Company, 26, 621–634. <https://doi.org/10.1002/ajim.4700260505>.

Rosenstock, L. (ed.) (1998), *What you need to know about occupational exposure to metalworking fluids*, Department of Health; Human Services (NIOSH).

Wen, L., Young, J. G., Robins, J. M., and Hernán, M. A. (2020), “Parametric g-formula implementations for causal survival analyses,” *Biometrics*, Wiley Online Library.

Woskie, S. R., Smith, T. J., Hallock, M. F., Hammond, S. K., Rosenthal, F., Eisen, E. A., Kriebel, D., and Greaves, I. A. (1994), “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*, Taylor & Francis, 55, 20–29.

Woskie, S. R., Virji, M. A., Hallock, M., Smith, T. J., and Hammond, S. K. (2003), “Summary of the findings from the exposure assessments for metalworking fluid mortality and morbidity studies,” *Applied occupational and environmental hygiene*, Taylor & Francis, 18, 855–864.