Application to the UAW-GM Cohort

In the simulation study, we showed that under several scenarios compatible with our hypothetical causal structure, the AWKM survival estimator had small bias compared to the WKM estimator. The bias was smaller when the cumulative incidence of the outcome was low and at later follow-up time points. Next, we estimated cancer-free survival in a real-world context. Using data from two plants participating in the UAW-GM Cohort study, we followed 38 553 individuals starting from hire to 40 years after hire for incidence of digestive system cancers (colon, rectal, esophageal, or stomach). This duration of follow-up spans the vast majority of individuals' working lifetimes. As in the simulation, the UAW-GM data were longitudinal data with baseline covariates, time-varying covariates, and a survival outcome.

The exposures of interest were MWF of three types: straight, soluble, and synthetic (Byers 2006; F. Mirer 2003; F. E. Mirer 2010). Straight MWFs are hydrocarbon-based fluids whose use became widespread by the 1920s. They continue to be widely used because of their simple formulation. In straight MWFs, hydrocarbons of different lengths are mixed together with other additives to attain different properties. Straight MWFs contain polycyclic aromatic hydrocarbons, long known to be carcinogenic (IARC 1973). Soluble oils are water-based oil emulsions first introduced in response to rising oil prices. They now make up the largest market share of MWFs (Childers 2006). Soluble oils are vulnerable to microbial contamination, so the addition of biocides is needed. However, their high lubricity makes soluble MWFs the most popular fluid type. Synthetic MWFs have the best toxicological profile, have no oil, and have a higher resistance to microbial growth. They were introduced into the MWF market in the second half of the 20th C., but fail to out-perform soluble MWFs in metalworking applications. Soluble and synthetic MWFs contain biocides, corrosion inhibitors, and chlorinated compounds, some classified as carcinogenic by the IARC (IARC 1987).

The outcome of interest was digestive system cancer incidence. There is little past research linking digestive system cancers to MWF exposures, but there is some evidence suggesting that straight MWFs cause digestive system cancers (Izano et al. 2019). Cancer incidence was obtained by linkage to Surveillance, Epidemiology, and End Results (SEER), which recorded cancer incidence cases starting on January 1, 1973. The cohort is comprised of individuals hired between 1938 and 1975. Cancer-free survival to the start of the registry was a left-filtering process possibly in the presence of the HWSE as investigated in the simulations. Over the 40 year follow-up period, vital status was obtained through the Social Security Administration, the National Death Index, as well as records provided by the UAW. The exposure rules of interest in the applied analysis were different than those in the simulation study: a_0 being exposed with 75% chance in the years exposure was observed and a_1 being exposed in the years exposure was observed. Weights were truncated at 1000. Binary exposure is defined to be exposure above the median level of exposure to straight, soluble, and synthetic MWFs or not. Counterfactual survival under rules a_0 and a_1 were estimated using the WKM and AWKM estimators. Baseline confounders included race, sex, plant, and year of hire. Time-varying confounders included age, cumulative time off, employment status, and exposure to the metalworking fluids exposure from the past year. These terms were included in the estimation of both the treatment and censoring mechanisms, which were estimated with stratification on every two years of follow-up. Summary statistics for the full study population and the digestive system cancer cases are presented in Table 1.

Table 1: Study population characteristics.

	Full cohort		Digestive cancer cases	
n (person-years)	26 182	(695 475)	213	(6000)
Race (%)				
Black	6 017	(23.0)	66	(31.0)
White	20 165	(77.0)	147	(69.0)
Sex (%)				
Female	3 328	(12.7)	15	(7.0)
Male	22 854	(87.3)	198	(93.0)
Plant (%)				
Plant 1	9 092	(34.7)	103	(48.4)
Plant 2	17 090	(65.3)	110	(51.6)
Ever exposed to MWF (%)	13 240	(50.6)	95	(44.6)
Year of hire (mean (SD))	1963	(12.26)	1960	(9.55)
Age at end of follow-up (mean (SD))	55.09	(12.02)	63.58	(9.40)
Cumulative years off (mean (SD))	0.06	(0.15)	0.12	(0.24)

Assumptions

Since we are working with observational data, the evaluation of the no-interference, causal consistency, ignorability, and overlap (positivity) assumptions are critical for causal inference. The stability of our estimation depends on positivity, which we assessed qualitatively by examining the distribution of the weights. The no-interference assumption may be problematized by the fact that there were a finite number of job types in the factory setting. If one worker operates a particular metalworking machine, then the the other workers are not able to operate that machine at that time. Instead, they may be assigned to assembly tasks, which have lower MWF exposure opportunities. That said, since these factories were quite large, there may be approximate independence. The consistency assumption is also problematic. The MWFs of interest are complex chemical mixtures whose composition changes by design and by nature. Over the last several decades, the formulation of MWFs has changed significantly in reaction to performance needs and toxicity concerns (F. Mirer 2003; Byers 2006). The composition of MWFs also undergoes unintentional changes because of the nature of their use: MWFs are often applied in contexts where contamination by other substances and microbes is possible and chemical changes due to heat and pressure are likely. Indeed, there are substantial concerns over the carcinogenicity of the chemical species formed in the MWF mixtures that were not originally added (Hidajat et al. 2020). Concerns regarding the consistency assumption may be abrogated in part by adequate adjustment for secular and factory-level characteristics.

Another key assumption meriting discussion is that of conditional ignorability. In order to achieve identification, even in the absence of left filtering, we need to have conditionally ignorable future exposure status and ignorable future censoring status at each time point given past data. In occupational cohorts, employment status and health history are strong predictors of future death (Häfner 1987; Halliday 2014; Laliotis and Stavropoulou 2018). Logically, major causes of death first act through employment status before they precipitate death. This dynamic is actually a key component in the setup for HWSE. We are therefore relatively confident that conditional ignorability of censoring due to death is attained given covariate, exposure, and cancer history. Our confidence in the conditional ignorability of exposure given history was not as strong. In particular, workers may be assigned to certain tasks based on their specific skills and knowledge, which may be associated with structural privileges that confer a lower risk of deleterious health outcomes. The potential magnitude of this uncontrolled confounding may have been bounded, however. The education level of the workers in the cohort was homogeneous, and all cohort members were members of the UAW union, which had uniform procedures in place for equitable access to training, wages, and career advancement (Harbison 1950; Barnard 2005). The presence of UAW policies support the assertion that given time since hire, job types (and therefore exposures) were randomly allocated.

Results

Estimated digestive system cancer-free curves under rules a_0 and a_1 applied to the three MWF types are presented in Figure **??**. The survival curves are more or less overlapping under the two exposure rules. Numeric values for the cancer-free survival and difference in cancer-free survival are presented in Tables **??** and **??**. The numeric summaries are consistent with the qualitative interpretation of the estimated survival curves; for all MWF types, expected survival time does not differ substantially across rules a_0 and a_1.

Figure **??** presents the median cumulative weight applied years since hire with ribbons showing the minimum and maximum weight. The overlap in the distribution of weights under the two rules for synthetic MWF exposure suggests that the model for the treatment and censoring mechanisms were inadequate in distinguishing the units of analysis by their probability of following a certain exposure rule or of remaining alive. The distribution of the weights was very skewed. Without truncation, they would have been in the order of magnitude of 10^20 or higher. Distributional summaries of the cumulative weights without truncation are presented in Table **??**. The presence of extremely large weights suggests that our observed data were inadequate for answering the causal question of interest due to practical violations in overlap (Petersen et al. 2012).

Barnard, John. 2005. *American Vanguard: The United Auto Workers During the Reuther Years, 1935-1970.* Wayne State University Press.

Byers, Jerry P. 2006. *Metalworking Fluids.* CRC Press.

Childers, Jean. 2006. "The Chemistry of Metalworking Fluids." In *Metalworking Fluids.* CRC Press.

Halliday, Timothy J. 2014. "Unemployment and Mortality: Evidence from the PSID." *Soc Sci Med* 113 (July): 15–22. https://doi.org/10.1016/j.socscimed.2014.04.038.

Harbison, Frederick H. 1950. "The General Motors-United Auto Workers Agreement of 1950." *Journal of Political Economy* 58 (5): 397–411.

Häfner, H. 1987. "Unemployment and Health." *Dtsch Med Wochenschr* 112 (37): 1428–32. https://doi.org/10.1055/s-2008-1068265.

Hidajat, Mira, Damien Martin McElvenny, Peter Ritchie, Andrew Darnton, William Mueller, Raymond M Agius, John W Cherrie, and Frank de Vocht. 2020. "Lifetime Cumulative Exposure to Rubber Dust, Fumes and n-Nitrosamines and Non-Cancer Mortality: A 49-Year Follow-up of UK Rubber Factory Workers." *Occup Environ Med* 77 (5): 316–23. https://doi.org/10.1136/oemed-2019-106269.

IARC. 1973. *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.

———. 1987. *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. Overall Evaluations of Carcinogenicity: An Updating of IARC Monographs, Volumes 1 to 42. World Health Organization.* Vol. 7. World Health Organization International Agency for Research on Cancer.

Izano, Monika A, Oleg A Sofrygin, Sally Picciotto, Patrick T Bradshaw, and Ellen A Eisen. 2019. "Metalworking Fluids and Colon Cancer Risk: Longitudinal Targeted Minimum Loss-Based Estimation." *Environmental Epidemiology* 3 (1): e035.

Laliotis, Ioannis, and Charitini Stavropoulou. 2018. "Crises and Mortality: Does the Level of Unemployment Matter?" *Soc Sci Med* 214 (October): 99–109. https://doi.org/10.1016/j.socscimed.2018.08.016.

Mirer, Franklin. 2003. "Updated Epidemiology of Workers Exposed to Metalworking Fluids Provides Sufficient Evidence for Carcinogenicity." *Applied Occupational and Environmental Hygiene* 18 (11): 902–12.

Mirer, Franklin E. 2010. "New Evidence on the Health Hazards and Control of Metalworking Fluids Since Completion of the OSHA Advisory Committee Report." *American Journal of Industrial Medicine* 53 (8): 792–801.

Petersen, Maya L, Kristin E Porter, Susan Gruber, Yue Wang, and Mark J van der Laan. 2012. "Diagnosing and Responding to Violations in the Positivity Assumption." *Statistical Methods in Medical Research* 21 (1): 31–54. https://doi.org/10.1177/0962280210386207.