Risk of non-Hodgkin lymphoma under hypothetical interventions with guaranteed positivity

Kevin T. Chen[1], Sally Picciotto[2], Ellen A. Eisen[2]

University of California, Berkeley – School of Public Health

Backgound

- **Non-Hodgkin Lymphoma** (NHL) incidence was associated with exposure to **soluble metalworking fluid** (MWF) in a Cox analysis of the **United Auto Workers-General Motors** (UAW-GM) **cohort**
- Unlike traditional regression analysis, *causal inference* methods
  - Can adjust for time-varying confounding affected by past exposure
  - Estimate population effects of hypothetical interventions
- Causal inference in statistics requires **positivity**, which is **not always assessed or addressed**
- Here, we specified ***supportable* interventions** on soluble MWF exposure in the UAW-GM cohort that **guarantee positivity**
- We estimated the effect of supportable interventions on NHL risk in the UAW-GM cohort (1985-2015) using the **hazard-extended iterative conditional expectation (ICE) parametric g-formula**
- Unlike the classic parametric g-formula, ICE g-formula estimators do not require parametric specification of the full joint distribution of the confounders, exposure, and outcome

Target and supportable exposure limits

- Suppose we want to know the effect of capping exposures at a **target exposure limit**, but estimation may not be supported by observed data
- Instead, we define **supportable exposure limits** for every time k and unique combination of confounder and exposure histories $(\bar{l}_k, \bar{a}_{k-1})$
  - Limit at greatest observed exposure $\leq$ **target limit**, if exists
  - No limit, if all observed exposures $>$ **target limit**
- Propensity scores for exposure at the supportable exposure limits are guaranteed to be **strictly positive**
- **Supportable intervention rule** for every $(\bar{l}_k, \bar{a}_{k-1})$, reduces exposures a_k above the **supportable exposure limit** to that limit, but allows exposures below to vary naturally (Figures 1 and 2)
- Applying the supportable intervention rule to the observed data induces the intervention distribution, which defines the **stochastic dynamic intervention** with **guaranteed positvity**

Figure 1: Marginal distribution of nonzero exposure at $k = 2$ before and after applying the supportable intervention rule

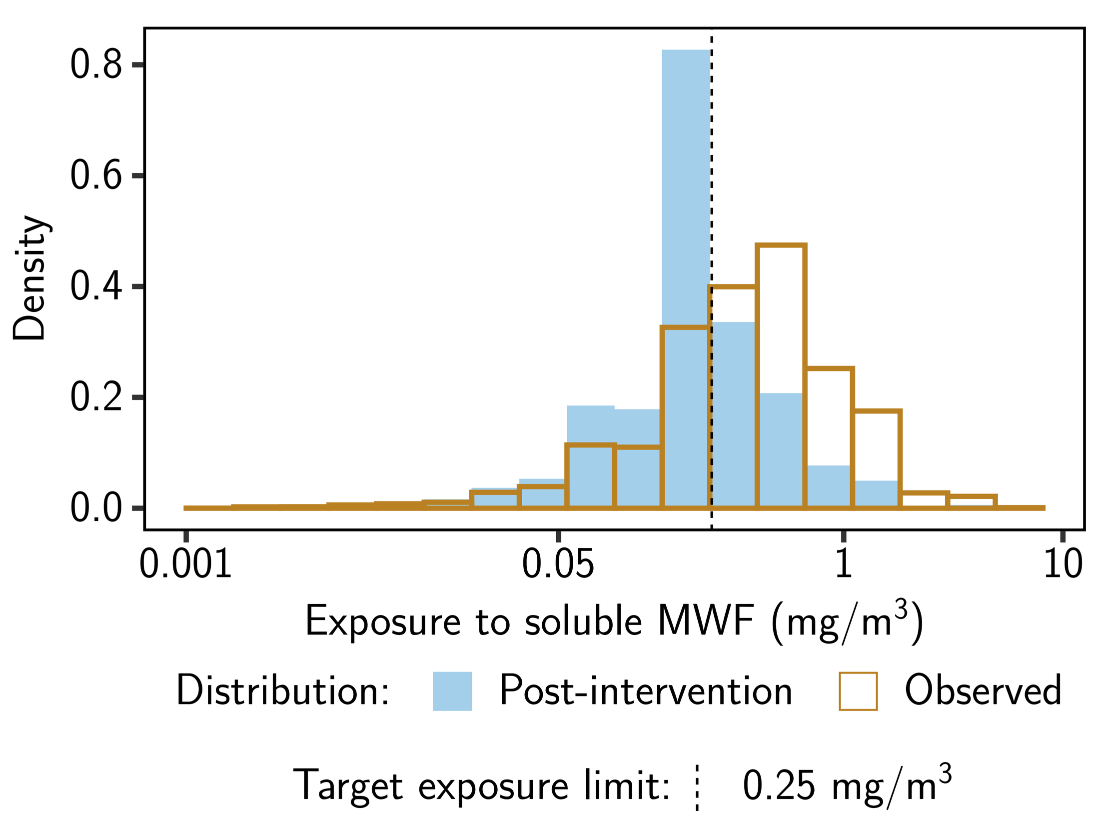


Figure 2: Distribution of nonzero exposure for three distinct confounder and exposure histories at $k = 2$ before and after applying the supportable intervention rule

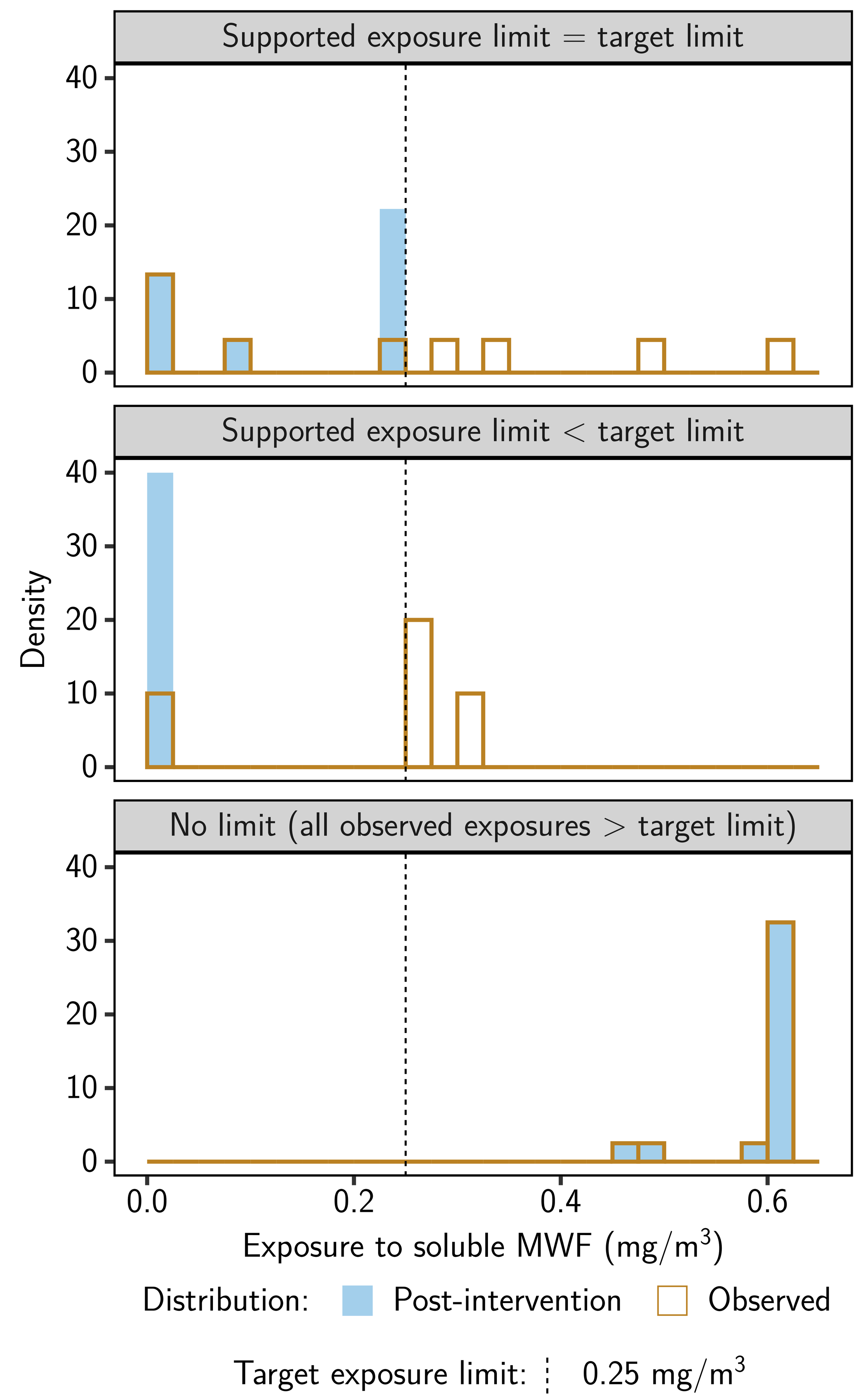


Table 1: Demographic characteristics of the full cohort and the NHL cases

	Study population	NHL cases
N (person-years)	33,134 (794,733)	339 (5,809)
Race		
White	21,315 (64%)	250 (74%)
Black	6,250 (19%)	40 (12%)
Unknown	5,569 (17%)	49 (14%)
Sex		
Male	30,249 (87%)	206 (89%)
Female	4,499 (13%)	25 (11%)
Years at work	15.2 (7.0, 26.6)	21.0 (7.8, 29.9)
Cumulative exposure$^\flat$	4.33 (1.71, 10.69)	5.43 (2.19, 14.33)

Statistics shown are count (percent) or median (first and third quartiles).
$^\flat$ Among those who were ever-exposed; units in mg/m^3 · year.

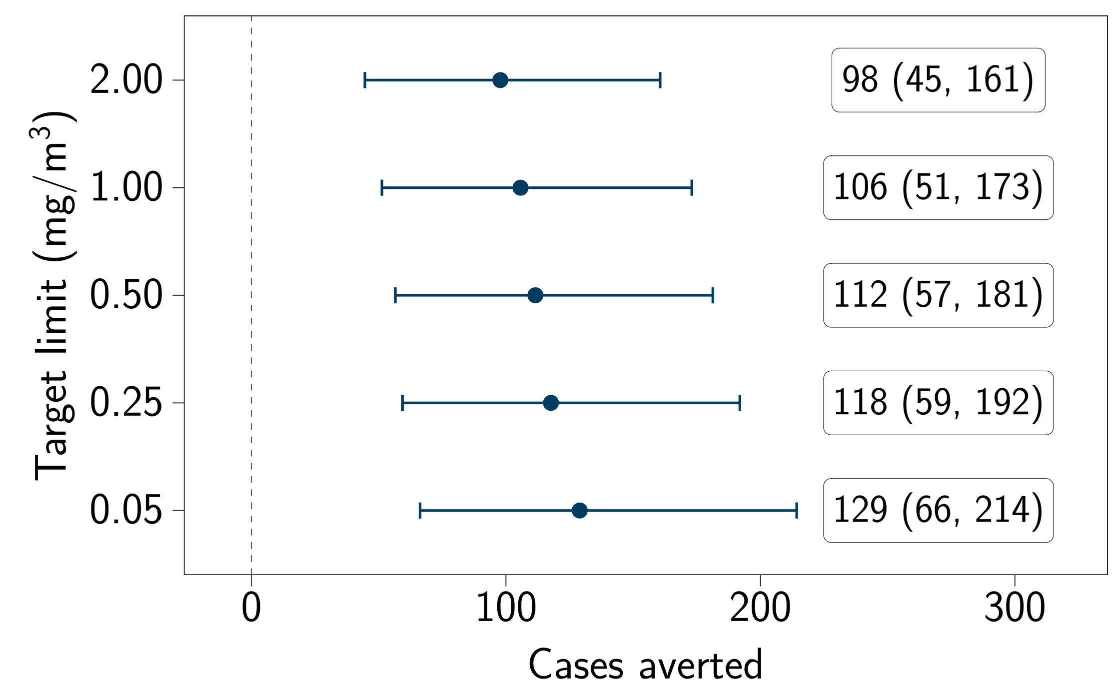
Affiliations and acknowledgements

[1] Division of Epidemiology & Biostatistics [2] Division of Environmental Health Sciences

Funding: CDC, NIOSH (R01OH011092) and NIOSH Training Grant (T42OH008429)

Main results

Figure 3: Counterfactual number of cases averted under supportable intervention rules based on five different target exposure limits and no censoring, with 95% bootstrap confidence intervals.



- There would have been **502 NHL cases** if there were no censoring
- Stronger target exposure limits **monotonically reduced** NHL risk
- Setting the target exposure limit at the **NIOSH recommended exposure limit** 0.5 mg/m^3 for total particulate mass derived from MWF **would have averted 112 (95% CI: 57, 181) NHL cases**

Conclusions and discussion

- **Stronger limits on exposure to soluble MWF provide stronger protections against NHL**
- During the anticipated rebound in domestic manufacturing, protecting worker health should be a priority
- We evaluated supportable interventions with **guaranteed positivity** and therefore **avoid extrapolation**
- We expect uniformly-enforced target exposure limits to have even **stronger protective effects**
- The **classic parametric g-formula** estimator can also estimate effects of supportable intervention rules, but **requires many more parametric assumptions** than ICE g-formula estimators

References

[1] Hilary L Colbeth et al. "Exposure to Metalworking Fluids and Cancer Incidence in the United Auto Workers–General Motors Cohort". In: *American Journal of Epidemiology* (2022).

[2] H. Michael Arrighi and Irva Hertz-Picciotto. "The evolving concept of the healthy worker survivor effect". In: *Epidemiology* (1994).

[3] Maya L Petersen et al. "Diagnosing and responding to violations in the positivity assumption". In: *Statistical Methods in Medical Research* (2012).

[4] Lan Wen et al. "Parametric g-formula implementations for causal survival analyses". In: *Biometrics* (2020).

[5] Jessica G Young, Miguel A Hernán, and James M Robins. "Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data". In: *Epidemiologic methods* (2014).

[6] Sarah L Taubman et al. "Intervening on risk factors for coronary heart disease: an application of the parametric g-formula". In: *International journal of epidemiology* (2009).