GM-UAW Cohort Study

"Weighted'' cumulative incidence

November 5, 2022

1. Use pooled logistic regression to approximate the $J = 2$ subdistribution hazard functions given covariates W and exposure X, where $j = 1$ indicates cancer incidence and $j = 2$ indicates natural cause mortality.

$$\begin{aligned} h_j(t \mid W, X) &= \lim_{\delta \to 0} \frac{1}{\delta} \mathbb{P}\left(t < T < t + \delta, J = j \mid (T > T \cup (T < t \cap J \neq j)), W, X\right) \\ &\approx \mathbb{P}\left(t < T < t + 1, J = j \mid (T > T \cup (T < t \cap J \neq j)), W, X\right) \equiv p_j(t \mid W, X) \end{aligned}$$

For natural cause mortality ($j = 2$), we start follow-up in the minmum of 1941 or three years after hire, including all who were hired after 1938. For cancer incidence $j = 1$, we start follow-up in 1973 (or 1985 if we do not use SEER). There is still censoring form loss to FU and death due to external causes.

2. For each person-year, evaluate $p_j(t \mid W = w, do(X = x))$ where x is the counterfactual exposure level of interest, and w is the observed covariate combination. Set $S_2(t \mid W, do(X = x)) = \prod^t (1 - p_2(t \mid W, do(X = x)))$. See Greenland's paper on model-based standardization.[1]

3. For each person-year starting in the first year of cancer incidence follow-up, take the product $p_1(t \mid w, do(X = x)) \cdot S_2(t \mid w, do(X = x))$.

4. Take the sum of products $\sum_t^T p_1(t \mid w, do(X = x)) \cdot S_2(t \mid w, do(X = x))$. This is a "weighted" estimate of the cumulative incidence.

5. Estimate cumulative incidence ratios by evaluating the sum in step 4 for different counterfactual levels of exposure x.

---

1. Greenland S. Model-based Estimation of Relative Risks and Other Epidemiologic Measures in Studies of Common Outcomes and in Case-Control Studies. *American Journal of Epidemiology*. 2004;160(4):301-305. doi:10.1093/aje/kwh221.