Estimating survival in left filtered data

Replication and application of Izano's estimator

Kevin Chen

Chapter 3 of Izano (2017)

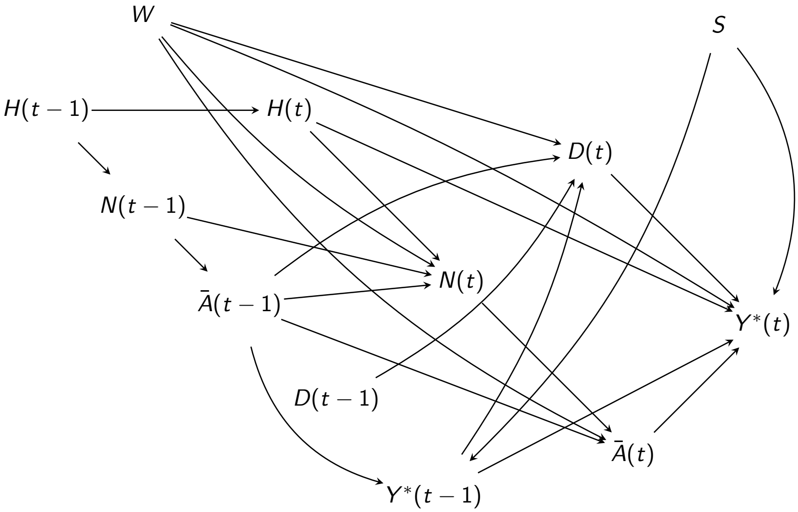
Drivers of Biased Effect Estimates in Left Filtered Data

- Specified a SEM for “left filtering” in the presence of HWSE
- Simulated large-sample ($n = 10^6$) data once for each of 5 scenarios
- Simulated finite-sample ($n = 50,000$) data 500 times for each scenario
- Applied two estimators:
    - Adjusted Kaplan-Meier
    - Delayed-entry adjusted Kaplan-Meier

Description of variables

Variable	Description
R	Time until start of registry
W	Baseline covariates
S	Susceptibility to effects of metalworking fluid exposure
$H(t)$	Adverse health status at time t
$N(t)$	Employment status at time t
$A(t)$	Metalworking fluid exposure at time t
$D(t)$	Mortality status at time t
$Y^*(t)$	Cancer status at time t
$Y(t)$	Observed Cancer status at time t
$t = \{1, 2, \ldots, 20\}$	Time, indexed in years after hire

Abbreviated DAG summarizing the causal structure



Scenarios

- Five scenarios represent 5 sets of parameters used to generate data according to the SEM
- Sets 2, 3, 4, and 5 differ from Set 1 by a single parameter each
  - Scenario 2: Greater cancer-related mortality
    (increase HR from 1.6 to 7.4)
  - Scenario 3: Greater proportion of susceptibles
    (increase proportion from 10% to 20%)
  - Scenario 4: Greater time-varying confounding by history of adverse health
    (decrease HR of being at work from 0.6 to 0.2
    and increase HR of cancer incidence from 2.0 to 5.5)
  - Scenario 5: Greater background incidence
    (increase baseline log hazard from -7 to -6)

Target and estimation

- Goal: estimate cancer-free survival indexed by time since hire
  - Under rule where all are exposed while at work
  - Under rule where all are not exposed (ever)
  - Competing risk of death is accounted for by preventing death
- Adjusted KM ie the inverse probability of treatment weighted KM (WKM)
- Delayed-entry adjusted KM ie the Aalen-filtered WKM (AWKM)

Review of the Kaplan-Meier Estimator

- Let $c(t)$ be the number of cases in the interval $(t-1, t]$
- Let $r(t)$ be the number of people at risk in interval $(t-1, t]$
- The standard survival estimator is

$$\widehat{S}(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{j \le t} \left(1 - \frac{c(j)}{r(j)}\right) & \text{if } t \ge t_1 \end{cases}$$

  where t_1 is the first event time
- Restricting to followers of rule a, we have

$$c_a(t), \quad r_a(t), \quad \widehat{S}_a(t)$$

Mathematical expression of $c_a(t)$ and $r_a(t)$

$$c_a(t) = \sum_i^n \underbrace{\mathbb{1}\left[Y_i(t) = 1\right]}_{\text{Cancer by time } t} \times \underbrace{\mathbb{1}\left[Y_i(t-1) = 0\right]}_{\text{At-risk at time } t-1} \times \underbrace{\mathbb{1}\left[\bar{A}_i(t) = \bar{a}(t)\right]}_{\text{Followed rule } a}$$

$$r_a(t) = \sum_i^n \underbrace{\mathbb{1}\left[Y_i(t-1) = 0\right]}_{\text{At-risk at time } t-1} \times \underbrace{\mathbb{1}\left[\bar{A}_i(t) = \bar{a}(t)\right]}_{\text{Followed rule } a}.$$

Inverse probability of treatment weighted KM (WKM)

The ith person at time t is weighted by $w_{i,a}(t)$ the predicted probability of following rule a through time t:

$$c_a^w(t) = \sum_i^n w_{i,a}(t) \times \mathbb{1}\left[Y_i(t) = 1\right] \times \mathbb{1}\left[Y_i(t-1) = 0\right] \times \mathbb{1}\left[\bar{A}_i(t) = \bar{a}(t)\right]$$
$$r_a^w(t) = \sum_i^n w_{i,a}(t) \times \mathbb{1}\left[Y_i(t-1) = 0\right] \times \mathbb{1}\left[\bar{A}_i(t) = \bar{a}(t)\right].$$

The $w_{i,a}$ can be estimated by fitting logistic regressions.

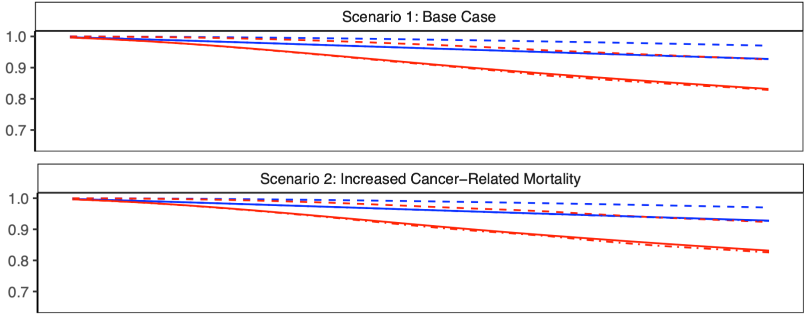
Aalen-filtered WKM (AWKM)

In addition to weighting, we restrict the computation of the discrete hazard to those under observation:

$$c_a(t) = \sum_i^n w_{i,a}(t) \times \mathbb{1}\left[Y_i(t) = 1\right] \times \mathbb{1}\left[Y_i(t-1) = 0\right] \times \mathbb{1}\left[\bar{A}_i(t) = \bar{a}(t)\right] \times \mathbb{1}\left[t \geq R_i\right]$$

$$r_a(t) = \sum_i^n w_{i,a}(t) \times \mathbb{1}\left[Y_i(t-1) = 0\right] \times \mathbb{1}\left[\bar{A}_i(t) = \bar{a}(t)\right] \times \mathbb{1}\left[t \geq R_i\right]$$

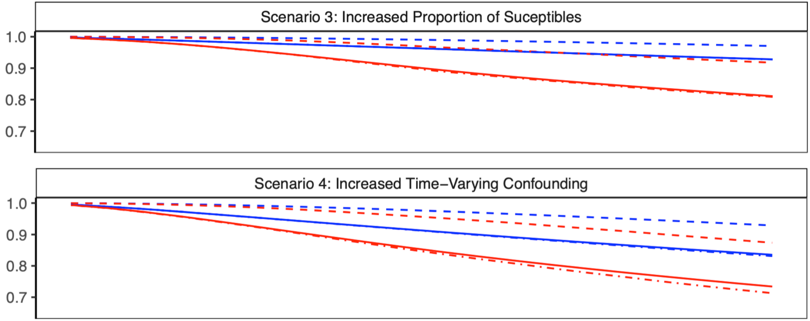
Results from Izano's Dissertation (2017)



Red: Always exposed   Blue: Never exposed

Long dashed: WKM   Dot dashed: AWKM   Solid: True KM curve

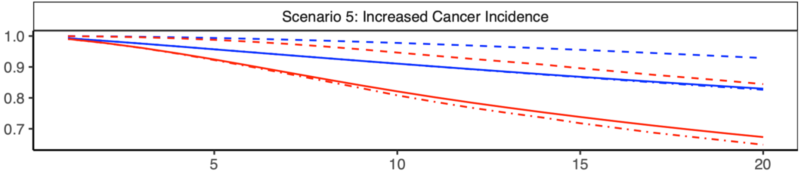
Results from Izano's Dissertation (2017)



Red: Always exposed Blue: Never exposed

Long dashed: WKM Dot dashed: AWKM Solid: True KM curve

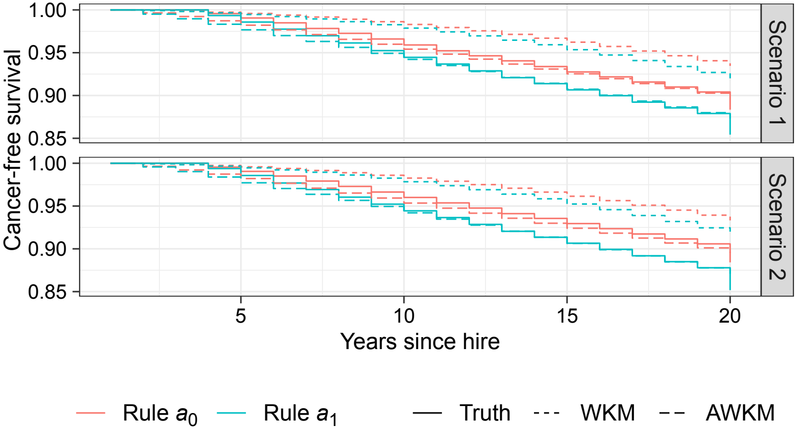
Results from Izano's Dissertation (2017)



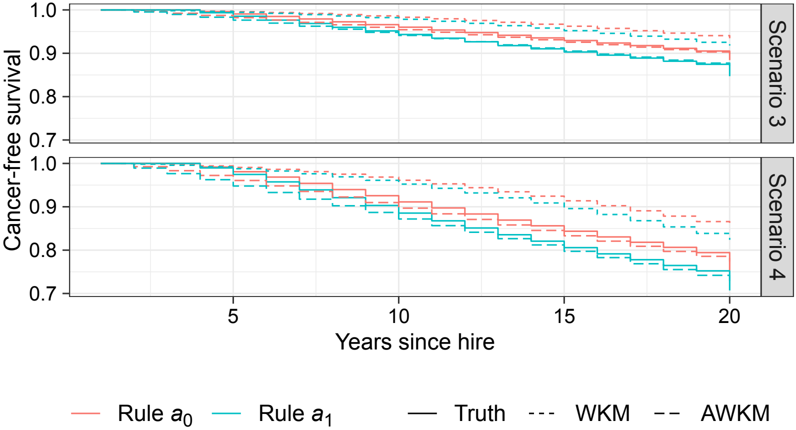
Red: Always exposed   Blue: Never exposed

Long dashed: WKM   Dot dashed: AWKM   Solid: True KM curve

Replication results



Replication results



Replication results

Application to UAW-GM

- Study population: Workers from Plants 1 or 2, hired after 1938
- 55-year follow-up starting at hire
- MWF exposure lagged 15 years
- Employment records end in 1994; workers considered administratively censored if they were still at work in 2010 (15 years after1 995)
- Rules of interest
    - Natural course: ever-exposed above reference level (0 $mg/m^3 \cdot$years for straights and synthetics; 0.05 $mg/m^3 \cdot$years for solubles)
    - Never exposed
- Under both rules, no censoring by death

Application results