Risk of incident prostate cancer risk, weighted by survival to 1985

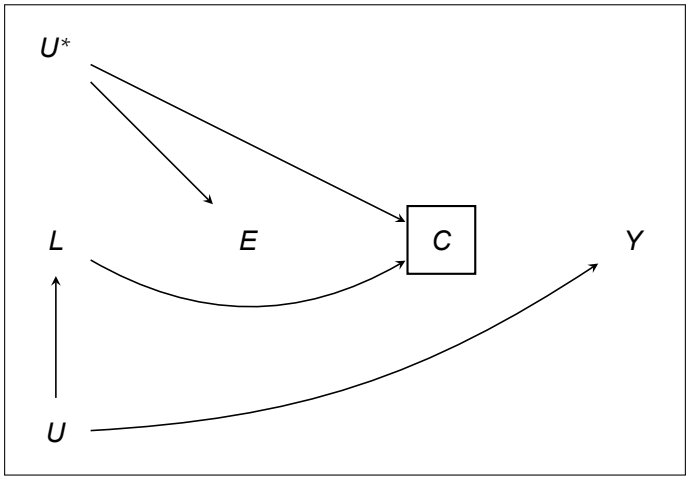
GM-UAW Cohort Study

June 26, 2020

The problem

The Cohort includes all persons who have worked at least 3 years and were hired in the years 1938 through 1981. However, follow-up for cancer incidence (Michigan Cancer Registry) does not begin until 1985. Not all who entered the cohort prior to 1985 were still at-risk of the outcome in 1985. Loss to follow-up may have been differential by underlying health. In particular, healthier individuals may have had lower exposure and a greater probability of surviving to 1985 outcome-free. In this case, the resulting population at-risk in 1985 would be the result of the so-called healthy worker survivor bias.[1] (There may also be those who have already experienced incident cancer prior to 1985, but were not recorded as such, but this outcome misclassification problem is not the focus of this analysis.)

Figure 1: Adapted from Figure 6c in Hernán et al. (2004). Take L, E, C, and Y to represent underlying health, occupational exposure, censoring, and cancer incidence, respectively.



Workaround

We have no measure of underlying health L, but we proceeded by modeling the censoring mechanism $\mathbb{P}_i(C_t \mid X_t, C_{t-1} = 0)$ for individual i at year t, conditional on observed covariates X_t that may serve as instruments for unmeasured L. We then used the modeled censoring probability $g_i(C_t \mid X_t, C_{t-1} = 0)$ to construct stabilized weights sw_i for each

individual in $t = 1985$ where

$$sw_i = \frac{1 - \frac{1}{n}\sum_i^n \prod_t^{1984} g_i(C_t = 0 \mid X_t, C_{(t-1)} = 0)}{1 - \prod_t^{1984} g_i(C_t = 0 \mid X_t, C_{(t-1)} = 0)}.$$

The parameters of model $g_i(C_t \mid X_t, C_{t-1} = 0)$ were estimated using a pooled logistic regression for the log-odds of dying due to natural causes by the end of each year of follow-up $t = \{1, \ldots, T\}$, conditional on the P-length covariate vector $X_t = x_t$ and uncensored status at the begninnig of that person-year $C_{t-1} = 0$.

$$\text{log}\, \frac{\hat{\mathbb{P}}\left(C_t = 1 \mid X_t = x_t, \hat{\beta}, C_{t-1} = 0\right)}{1 - \hat{\mathbb{P}}\left(C_t = 1 \mid X_t = x_t, \hat{\beta}, C_{t-1} = 0\right)} = \hat{\beta}_0 + \hat{\beta}_1 X_{1t} + \cdots + \hat{\beta}_P X_{Pt}.$$

The covariates X_t included:

- Years since hire (quartiles or splined)
- Age (quartiles or splined)
- Plant
- Race (black or white)
- Sex
- Proportion of year spent in assembly, machining (includes grinding), and off (quartiles)
- Cumulative time spent off (quartiles)
- Year of hire (quartiles)
- Cumulative exposure to straight, soluble, and synthetic MWFs (quartiles)
- Employment status

Further information on the estimation of g can be found on Box.

Four analyses

1. Men still alive in 1973
2. Men still alive in 1985, excluding the 134 men who had prostate cancer before 1985
3. Men still alive in 1985, excluding the 134, weighted by probability of not dying of natural causes
4. Men still alive in 1985, excluding the 134, weighted by stabilized inverse probability with truncation at the 99th percentile of weights (sw_i = 330)

Follow-up ends on the date of death, prostate cancer incidence, LTFU (outcome-free at age 108.39 years), or 2015-12-31, whichever comes first.

All analyses restricted to plants 1 and 2

Summary of population characteristics in the slides to follow.

Men still alive in 1973		
Study population size (N)	20 996	100%
Race		
White	11 192	53 %
Black	4 694	22 %
Unknown	5 110	24 %
Plant[♮]		
Plant 1	7 834	37 %
Plant 2	13 162	63 %
Diagnosed with cancer by end of follow-up	4 517	22 %
Years of follow-up	36.65	(31.73, 45.19)
Years at work[*]	15.66	(7.01, 26.29)
Year of hire	1963	(1951, 1975)
Age at hire (years)	25	(21, 33)
Year of birth	1934	(1921, 1948)
Year of first cancer diagnosis	1997	(1989, 2006)
Age at first cancer diagnosis (years)	67	(59, 74)
Cumulative exposure[♯] to MWFs (mg/m^3·y)		
Straight	0.5	(0.18, 1.43)
Soluble	5.05	(2.32, 12.49)
Synthetic	0.4	(0.15, 1.27)

Men still alive in 1985		
Study population size (N)	18 342	100%
Race		
White	10 424	57 %
Black	4 174	23 %
Unknown	3 744	20 %
Plant[♮]		
Plant 1	6 737	37 %
Plant 2	11 605	63 %
Diagnosed with cancer by end of follow-up	3 983	22 %
Years of follow-up	38.18	(34.36, 46.3)
Years at work[*]	15.57	(6.87, 27.13)
Year of hire	1965	(1952, 1976)
Age at hire (years)	24	(21, 31)
Year of birth	1938	(1924, 1950)
Year of first cancer diagnosis	1999	(1992, 2007)
Age at first cancer diagnosis (years)	67	(60, 74)
Cumulative exposure[♯] to MWFs (mg/m^3·y)		
Straight	0.5	(0.18, 1.43)
Soluble	4.89	(2.27, 12.26)
Synthetic	0.39	(0.15, 1.26)

Men still alive in 1985		
Study population size (N)	18 342	100%
Race		
White	10 424	57 %
Black	4 174	23 %
Unknown	3 744	20 %
Plant[♮]		
Plant 1	6 737	37 %
Plant 2	11 605	63 %
Plant 3	>0.00	0 %
Diagnosed with cancer by end of follow-up	3 983	22 %
Years of follow-up	38.18	(34.36, 46.3)
Years at work[*]	15.57	(6.87, 27.13)
Year of hire	1965	(1952, 1976)
Age at hire (years)	24	(21, 31)
Year of birth	1938	(1924, 1950)
Year of first cancer diagnosis	1999	(1992, 2007)
Age at first cancer diagnosis (years)	67	(60, 74)
Cumulative exposure[♯] to MWFs (mg/m^3·y)		
Straight	0.5	(0.18, 1.43)
Soluble	4.89	(2.27, 12.26)
Synthetic	0.39	(0.15, 1.26)

Cox PH Models

- Cumulative exposure to straight, soluble, and synthetic MWFs (lagged 20 years)
- Splined calendar year
- Splined year of hire
- Race
- Plant
- Risk sets indexed by age

1. Men still alive in 1973

Covariate	level	n	HR	(95% CI)
Cumulative straight	0	753	1.00	
	>0 to 0.322	287	1.05	(0.89, 1.24)
	>0.322 to 1.33	287	1.02	(0.87, 1.20)
	>1.33	287	1.08	(0.93, 1.25)
Cumulative soluble	0 to 0.05	176	1.00	
	>0.05 to 3.78	471	1.13	(0.93, 1.36)
	>3.78 to 13.3	483	1.09	(0.90, 1.33)
	>13.3	484	1.29	(1.06, 1.58)
Cumulative synthetic	0	1094	1.00	
	>0 to 0.22	174	0.96	(0.78, 1.19)
	>0.22 to 1.05	173	0.94	(0.76, 1.15)
	>1.05	173	1.07	(0.88, 1.30)
Race	White	961	1.00	
	Black	653	2.40	(2.13, 2.69)
Plant	1	813	1.00	
	2	801	1.02	(0.87, 1.19)
P-spline of calendar year (df = 14.68)		1614		
P-spline of year of hire (df = 13.67)		1614		

2. Men still alive in 1985

Covariate	level	n	HR	(95% CI)
Cumulative straight	0	672	1.00	
	>0 to 0.322	274	1.08	(0.91, 1.28)
	>0.322 to 1.33	261	1.03	(0.87, 1.22)
	>1.33	273	1.11	(0.96, 1.30)
Cumulative soluble	0 to 0.05	155	1.00	
	>0.05 to 3.78	449	1.10	(0.90, 1.34)
	>3.78 to 13.3	428	1.03	(0.84, 1.26)
	>13.3	448	1.25	(1.01, 1.55)
Cumulative synthetic	0	990	1.00	
	>0 to 0.22	171	0.99	(0.80, 1.23)
	>0.22 to 1.05	160	0.93	(0.75, 1.16)
	>1.05	159	1.05	(0.86, 1.29)
Race	White	875	1.00	
	Black	605	2.40	(2.12, 2.70)
Plant	1	745	1.00	
	2	735	1.01	(0.85, 1.19)
P-spline of calendar year (df = 16.99)		1480		
P-spline of year of hire (df = 14.85)		1480		

3. Men still alive in 1985, weighted by probability of not being dead due to natural causes in 1985

Covariate	level	n	HR	(95% CI)
Cumulative straight	0	672	1.00	
	>0 to 0.322	274	1.09	(0.92, 1.30)
	>0.322 to 1.33	261	1.06	(0.89, 1.25)
	>1.33	273	1.10	(0.95, 1.29)
Cumulative soluble	0 to 0.05	155	1.00	
	>0.05 to 3.78	449	1.10	(0.90, 1.35)
	>3.78 to 13.3	428	1.01	(0.82, 1.25)
	>13.3	448	1.24	(1.00, 1.54)
Cumulative synthetic	0	990	1.00	
	>0 to 0.22	171	0.97	(0.78, 1.22)
	>0.22 to 1.05	160	0.94	(0.76, 1.16)
	>1.05	159	1.05	(0.85, 1.28)
Race	White	875	1.00	
	Black	605	2.43	(2.14, 2.75)
Plant	1	745	1.00	
	2	735	0.98	(0.83, 1.16)
P-spline of calendar year (df = 16.99)		1480		
P-spline of year of hire (df = 14.78)		1480		

4. Men still alive in 1985, weighted by the stabilized inverse probability with truncation

Covariate	level	n	HR	(95% CI)
Cumulative straight	0	672	1.00	
	>0 to 0.322	274	1.64	(1.12, 2.41)
	>0.322 to 1.33	261	1.39	(0.94, 2.06)
	>1.33	273	1.23	(0.81, 1.87)
Cumulative soluble	0 to 0.05	155	1.00	
	>0.05 to 3.78	449	1.50	(0.88, 2.55)
	>3.78 to 13.3	428	1.51	(0.83, 2.74)
	>13.3	448	1.80	(0.99, 3.29)
Cumulative synthetic	0	990	1.00	
	>0 to 0.22	171	0.71	(0.43, 1.19)
	>0.22 to 1.05	160	0.66	(0.40, 1.07)
	>1.05	159	0.71	(0.43, 1.18)
Race	White	875	1.00	
	Black	605	2.25	(1.64, 3.10)
Plant	1	745	1.00	
	2	735	1.04	(0.66, 1.61)
P-spline of calendar year (df = 16.98)		1480		
P-spline of year of hire (df = 15.71)		1480		

Results by MWF type

Cumulative straight	1973		1985		1985, weighted		1985, s. weight	
	n	HR	n	HR	n	HR	n	HR
0	753	1.00	672	1.00	672	1.00	672	1.00
>0 to 0.322	287	1.05	274	1.08	274	1.09	274	1.64
>0.322 to 1.33	287	1.02	261	1.03	261	1.06	261	1.39
>1.33	287	1.08	273	1.11	273	1.10	273	1.23

Cumulative soluble	1973		1985		1985, weighted		1985, s. weight	
	n	HR	n	HR	n	HR	n	HR
0 to 0.05	176	1.00	155	1.00	155	1.00	155	1.00
>0.05 to 3.78	471	1.13	449	1.10	449	1.10	449	1.50
>3.78 to 13.3	483	1.09	428	1.03	428	1.01	428	1.51
>13.3	484	1.29	448	1.25	448	1.24	448	1.80

Cumulative synthetic	1973		1985		1985, weighted		1985, s. weight	
	n	HR	n	HR	n	HR	n	HR
0	1094	1.00	990	1.00	990	1.00	990	1.00
>0 to 0.22	174	0.96	171	0.99	171	0.97	171	0.71
>0.22 to 1.05	173	0.94	160	0.93	160	0.94	160	0.66
>1.05	173	1.07	159	1.05	159	1.05	159	0.71

1. Hernán MA, Hernández-Dı'az S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004:615-625.