UAW-GM Cohort Study

Predicting survival to 1985

June 1, 2020

Population

- Restricted to those:
  - Still alive in 1970
  - Hired in or after 1938, but no later than 1982
  - Missing no more than half of their work record
- Individuals contributed person-time from three years after hire or 1970 (whichever came first) to death or loss to follow-up
- Individuals were considered lost to follow-up upon reaching the oldest observed age at death (106.56 years)
- N = 36 986, (421 436 person-years)
- Deaths due to natural causes by end of 1984: 3 196 (8.6%)

ICD codes for natural causes of death

- ICD-9: all codes codes in [001, 799]
  - Excludes the categories labeled as "Injury and poisoning" and "external causes of injury and supplemental classification."
- ICD-10: all codes, except those with prefix S, T, V, W, X, or Y.

Pooled logistic regression

We use the pooled logistic regression model to estimate 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 covariates vector $\boldsymbol{X_t} = \boldsymbol{x}_t$ and aliveness at the begining of that person-year $Y_{t-1} = 0$.

$$\log \frac{\widehat{\mathbb{P}}\left(Y_t = 1 \mid \boldsymbol{X}_t = \boldsymbol{x}_t, \hat{\boldsymbol{\beta}}, Y_{t-1} = 0\right)}{1 - \widehat{\mathbb{P}}\left(Y_t = 1 \mid \boldsymbol{X}_t = \boldsymbol{x}_t, \hat{\boldsymbol{\beta}}, Y_{t-1} = 0\right)} = \hat{\beta}_0 + \hat{\beta}_1 X_{1t} + \cdots + \hat{\beta}_P X_{Pt}$$

where $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_P$ are the partial coefficient estimates.

Pooled logistic regression (continued)

- Covariates
  - 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
- Model 1: All covariates coded as categorical variables
- Model 2: Years since hire and age included in penalized splines

Model 1 results

Covariate	level	*n*	OR	95% CI
Intercept			0.00	(0.00, 0.00)
Years since hire	[4,8]	830		
	(8, 12]	842	0.99	(0.89, 1.09)
	(12, 16]	969	0.96	(0.87, 1.06)
	(16, 18]	555	0.96	(0.85, 1.08)
Age	[18.77,55.42]	799		
	(55.42, 62.84]	799	3.20	(2.86, 3.58)
	(62.84, 70.2]	799	4.20	(3.73, 4.72)
	(70.2, 106.6]	799	7.54	(6.65, 8.54)
Plant	1	1078		
	2	1625	0.81	(0.72, 0.90)
	3	493	0.61	(0.54, 0.70)
Race	White	2679		
	Black	517	0.90	(0.81, 1.01)
Sex	Male	3038		
	Female	158	0.42	(0.36, 0.50)
Time spent in assembly	0	2911		
	(0, 1]	285	1.10	(0.90, 1.33)
Time spent machining	0	2847		
	(0, 1]	349	0.79	(0.65, 0.98)

Model 1 results (continued)

Covariate	level	*n*	OR	95% CI
Time spent off	0	3084		
	(0, 1]	112	0.66	(0.53, 0.82)
Cumulative time off	0	2363		
	(0, 0.04932]	35	0.95	(0.67, 1.34)
	(0.04932, 6.402]	798	1.34	(1.22, 1.47)
Year of hire	[1938,1946]	799		
	(1946, 1951]	799	1.05	(0.95, 1.17)
	(1951, 1954]	801	1.10	(0.98, 1.22)
	(1954, 1982]	797	0.79	(0.70, 0.90)
Cumulative soluble exposure	[0,2.758]	799		
	(2.758, 8.944]	799	1.00	(0.90, 1.11)
	(8.944, 20.45]	799	1.12	(1.01, 1.25)
	(20.45, 240.8]	799	1.22	(1.09, 1.37)
Cumulative straight exposure	0	1309		
	(0, 0.1389]	289	1.03	(0.89, 1.18)
	(0.1389, 1.465]	799	1.05	(0.94, 1.16)
	(1.465, 293.4]	799	1.12	(1.01, 1.24)
Cumulative synthetic exposure	0	2173		
	(0, 0.2302]	224	0.90	(0.77, 1.06)
	(0.2302, 105]	799	1.07	(0.96, 1.18)
Employment status	At work	511		
	Left work	2685	3.12	(2.44, 3.97)

Model 2 results

Covariate	level	*n*	OR	95% CI
Intercept			0.00	(0.00, 0.00)
Plant	1	1078		
	2	1625	0.82	(0.74, 0.92)
	3	493	0.73	(0.64, 0.83)
Race	White	2679		
	Black	517	0.94	(0.84, 1.06)
Sex	Male	3038		
	Female	158	0.42	(0.35, 0.49)
Time spent in assembly	0	2911		
	(0, 1]	285	1.13	(0.93, 1.37)
Time spent machining	0	2847		
	(0, 1]	349	0.76	(0.62, 0.94)
Time spent off	0	3084		
	(0, 1]	112	0.73	(0.59, 0.91)
Cumulative time off	0	2363		
	(0, 0.04932]	35	1.10	(0.78, 1.55)
	(0.04932, 6.402]	798	1.49	(1.36, 1.64)

Model 2 results (continued)

Covariate	level	*n*	OR	95% CI
Year of hire	[1938,1946]	799		
	(1946, 1951]	799	1.08	(0.97, 1.20)
	(1951, 1954]	801	1.14	(1.02, 1.27)
	(1954, 1982]	797	1.23	(1.09, 1.38)
Cumulative soluble exposure	[0,2.758]	799		
	(2.758, 8.944]	799	0.91	(0.82, 1.01)
	(8.944, 20.45]	799	0.99	(0.89, 1.10)
	(20.45, 240.8]	799	1.09	(0.98, 1.22)
Cumulative straight exposure	0	1309		
	(0, 0.1389]	289	0.99	(0.86, 1.15)
	(0.1389, 1.465]	799	1.04	(0.93, 1.15)
	(1.465, 293.4]	799	1.09	(0.99, 1.20)
Cumulative synthetic exposure	0	2173		
	(0, 0.2302]	224	0.96	(0.82, 1.12)
	(0.2302, 105]	799	1.06	(0.96, 1.18)
Employment status	At work	511		
	Left work	2685	2.86	(2.25, 3.64)

Splined terms not shown.

From log-odds to survival

1. Extract the fitted probabilities $\hat{p}_{ti}$ from the pooled logistic regression (`fitted.values` of an object of class `lm`; the inverse of the link function has already been applied)
2. For each individual, with obserations ordered by time t, take the cumulative product $\prod^{t}(1 - \hat{p}_{ti})$
3. The cumulative product in each row represents the probability of survival (for natural cause mortality) to the end of that row's year

Average survival probabilities by race, sex and year of hire among those still alive in 1985

Covariate	level	Model 1	Model 2
Race	White	0.93	0.93
	Black	0.94	0.94
Sex	Men	0.93	0.93
	Women	0.97	0.97
Year of hire	[1938, 1946]	0.75	0.76
	(1946, 1951]	0.80	0.81
	(1951, 1954]	0.86	0.86
	(1954, 1982]	0.97	0.97

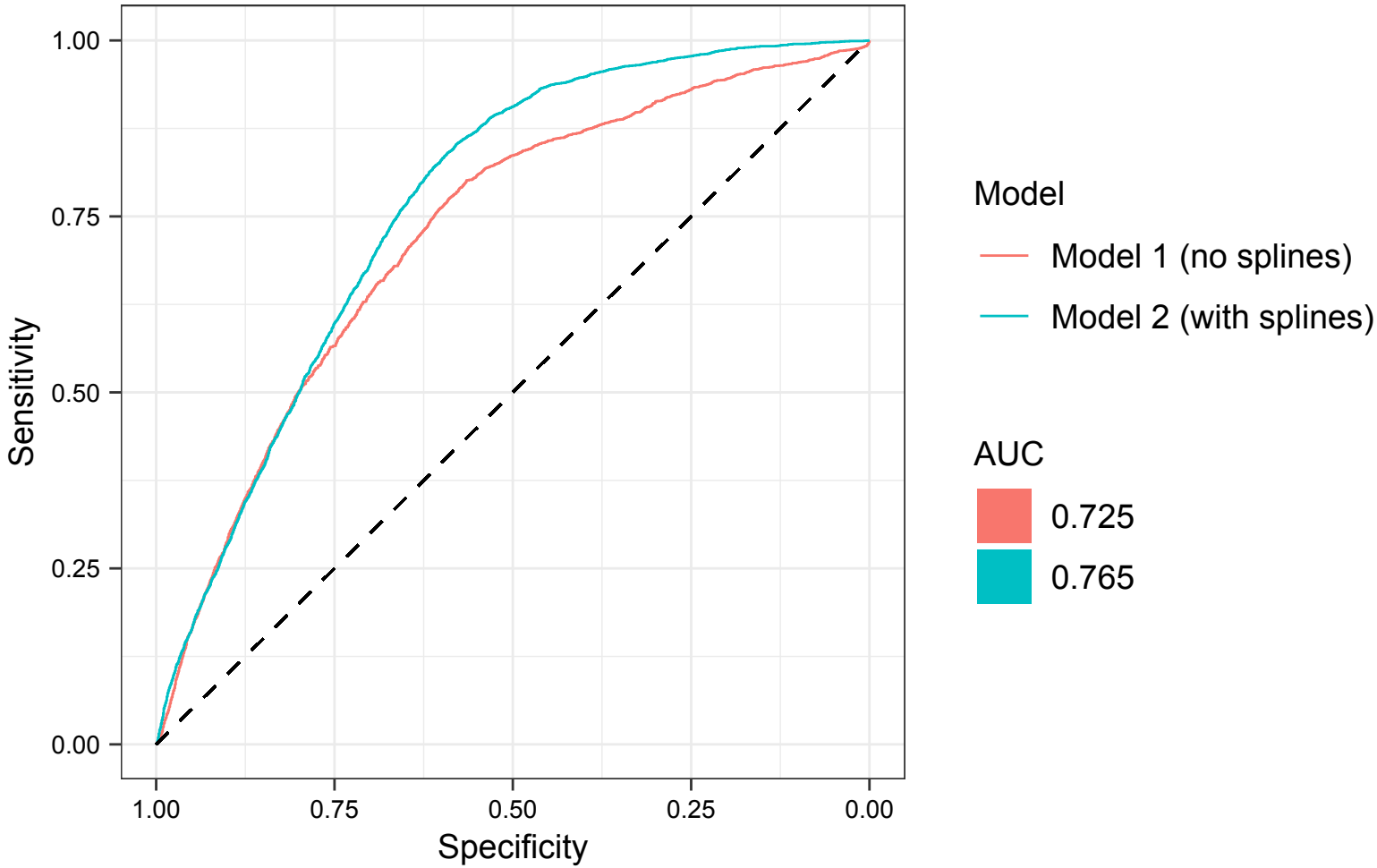
Average survival probabilities by age and employment status among those still alive in 1985

Covariate	level	Model 1	Model 2
Age	[18.77, 55.42]	0.98	0.99
	(55.42, 62.84]	0.90	0.89
	(62.84, 70.2]	0.82	0.81
	(70.2, 106.6]	0.64	0.65
Employment status	At work	0.98	0.99
	Left work	0.85	0.85

Average survival probabilities by MWF exposure among those still alive in 1985

Cumulative exposure	level	Model 1	Model 2
Straight	0	0.94	0.94
	(0, 0.1389]	0.94	0.94
	(0.1389, 1.465]	0.93	0.94
	(1.465, 293.4]	0.91	0.91
Soluble	[0, 2.758]	0.96	0.96
	(2.758, 8.944]	0.93	0.93
	(8.944, 20.45]	0.88	0.89
	(20.45, 240.8]	0.85	0.86
Synthetic	0	0.93	0.93
	(0, 0.2302]	0.96	0.97
	(0.2302, 105]	0.92	0.92

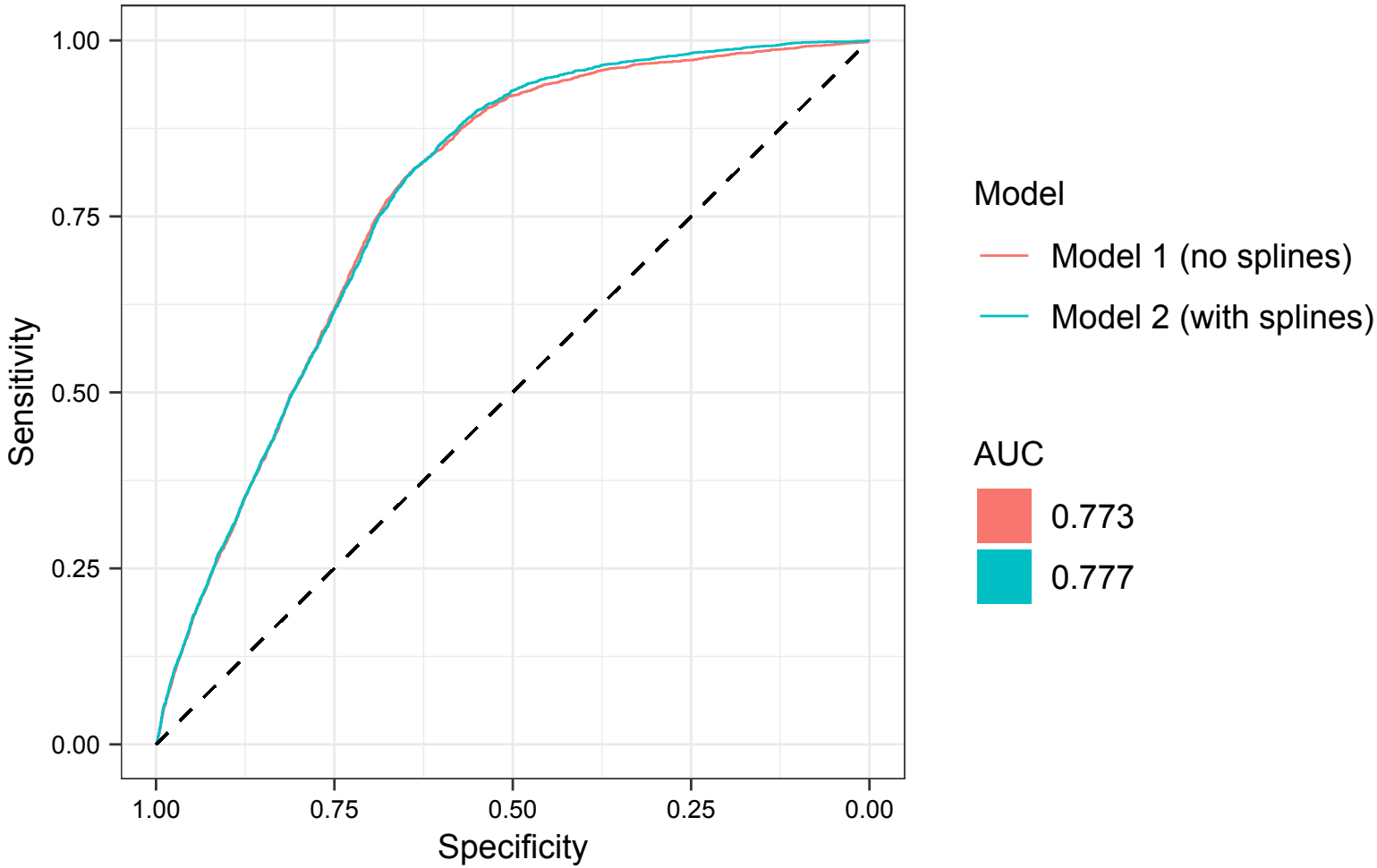
ROC Curve



Outcome: Natural cause mortality status in 1985

An individual's probability of not dying due to natural causes was calculated as: $\prod_t (1 - \hat{p}_t)$ where $\hat{p}_t$ is the predicted probability of death due to natural causes for the t^{th} year of follow-up.

Increase maximum number of levels from 4 to 20



Outcome: Natural cause mortality status in 1985

An individual's probability of not dying due to natural causes was calculated as: $\prod_t (1 - \hat{p}_t)$ where $\hat{p}_t$ is the predicted probability of death due to natural causes for the t^{th} year of follow-up.