SESSION 3: CHOICE OF THE TIME SCALE AND INTERACTIONS WITH TIME

Module 20: Survival Analysis for Observational Data

Summer Institute in Statistics for Clinical Research
University of Washington
July, 2017

Susanne May, Ph.D.

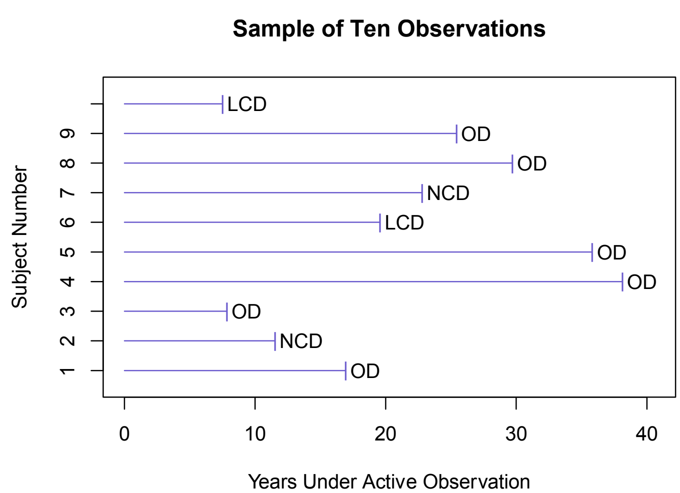
OUTLINE

- Choice of the time scale for analysis
- Left entry into observation (left truncation)
- Cox models including interaction with time variables
- Cox models with time-dependent coefficients

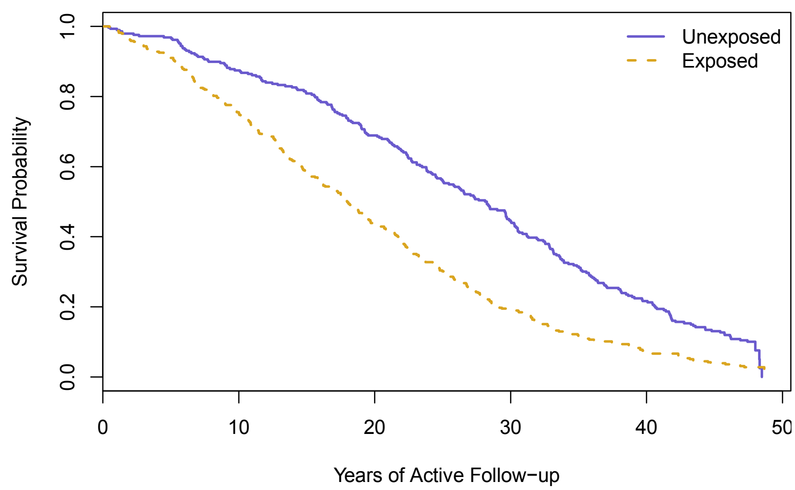
WELSH NICKEL REFINERS STUDY

- 679 nickel refinery workers identified twice on paysheets April 1929, 1934, 1939, 1944, 1949
- Follow-up until 1981
- Refinery cleaned up by various means 1922-1932, so all important exposure occurred before beginning of follow-up
- Interest in whether duration of employment in high-exposure areas, and age at first exposure were related to lung and nasal sinus cancer mortality risk.

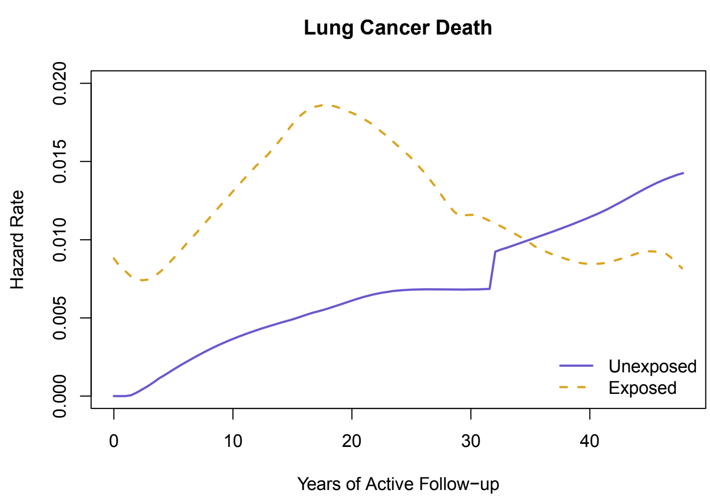
WELSH NICKEL REFINERS



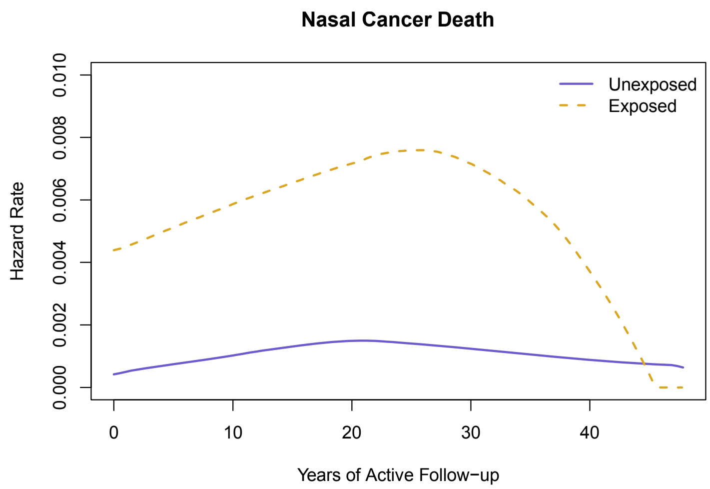
ALL-CAUSE MORTALITY



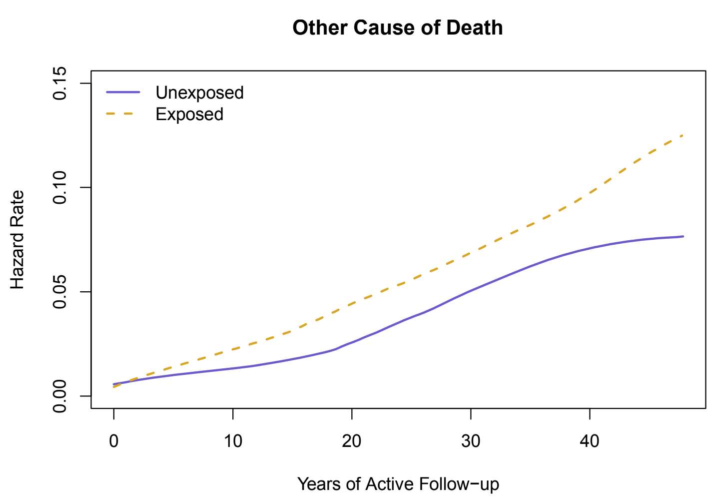
WELSH NICKEL REFINERS



WELSH NICKEL REFINERS



WELSH NICKEL REFINERS



LUNG CANCER FU TIME

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
exposedTRUE	0.9200182	2.509336	0.1869493	4.921217	9e-07

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
exp0.5 - 4.0	0.6030012	1.8275955	0.2121299	2.8426041	0.0044747
exp4.5 - 8.0	1.0862839	2.9632419	0.2828485	3.8405146	0.0001228
exp8.5-12.0	1.2772969	3.5869307	0.3742268	3.4131628	0.0006421
exp12.5+	1.4873597	4.4253955	0.4798472	3.0996524	0.0019375
afe20-27.5	0.8103938	2.2487934	0.3079688	2.6314149	0.0085030
afe27.5 - 35	0.9149895	2.4967489	0.3291081	2.7802097	0.0054324
afe35+	0.8068991	2.2409482	0.4237839	1.9040342	0.0569057
yfe1910-1914	0.3342204	1.3968510	0.2695145	1.2400835	0.2149445
yfe1915-1919	-0.1340505	0.8745459	0.3749097	-0.3575540	0.7206771
yfe1920-1925	0.0744977	1.0773429	0.2966621	0.2511197	0.8017216

NASAL CANCER FU TIME

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
exposedTRUE	1.614074	5.023236	0.3516507	4.589994	4.4e-06

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
exp0.5 - 4.0	0.8356274	2.3062606	0.4032111	2.072432	0.0382252
exp4.5 - 8.0	1.1366437	3.1162916	0.4706657	2.414970	0.0157365
exp8.5-12.0	2.2945326	9.9197981	0.5117936	4.483316	0.0000073
exp12.5+	2.8713357	17.6605917	0.5697217	5.039892	0.0000005
afe20-27.5	1.4686105	4.3431963	0.7518514	1.953326	0.0507810
afe27.5 - 35	2.1598639	8.6699580	0.7588726	2.846148	0.0044252
afe35+	3.4767227	32.3535148	0.7843101	4.432842	0.0000093
yfe1910-1914	0.7130093	2.0401213	0.3728470	1.912338	0.0558329
yfe1915-1919	0.5040978	1.6554913	0.5034466	1.001294	0.3166849
yfe1920-1925	-0.9304088	0.3943924	0.5152666	-1.805684	0.0709677

OTHER CAUSES FU TIME

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
exposedTRUE	0.3962896	1.4863	0.0972056	4.076818	4.57e-05

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
exp0.5 - 4.0	0.1318081	1.1408894	0.1105672	1.1921083	0.2332188
exp4.5 - 8.0	0.1308735	1.1398236	0.1603797	0.8160231	0.4144869
exp8.5-12.0	0.0324914	1.0330250	0.2563862	0.1267282	0.8991555
exp12.5+	-0.0774111	0.9255093	0.3964677	-0.1952520	0.8451957
afe20-27.5	0.5275548	1.6947832	0.1539622	3.4265217	0.0006114
afe27.5 - 35	1.1070376	3.0253827	0.1653356	6.6956992	0.0000000
afe35+	1.9740626	7.1998671	0.1942464	10.1626701	0.0000000
yfe1910-1914	-0.2148112	0.8066937	0.1515491	-1.4174361	0.1563555
yfe1915-1919	-0.5297679	0.5887416	0.1766843	-2.9983870	0.0027141
yfe1920-1925	-1.1456390	0.3180206	0.1502442	-7.6251795	0.0000000

COX REGRESSION MODEL

$$\lambda(t) = \lambda_0(t)e^{\beta_1 x_1 + \cdots + \beta_k x_k}$$

Interpretation of e^{β_1} in general:

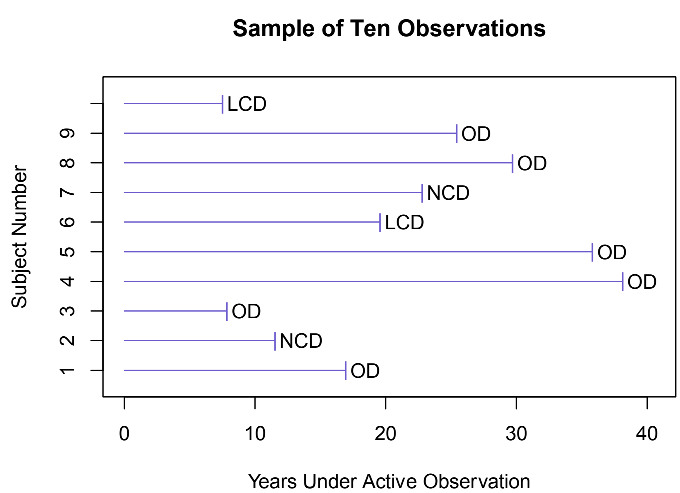
"Relative risk (or hazard ratio) associated with a one unit higher value of x_1, holding $x_2, \ldots, x_k$ constant".

$$\begin{aligned} \lambda(t) \text{ for } x_1 + 1: &\quad \lambda_0(t)e^{\beta_1(x_1+1) + \cdots + \beta_k x_k} \\ \lambda(t) \text{ for } x_1: &\quad \lambda_0(t)e^{\beta_1 x_1 + \cdots + \beta_k x_k} \\ \text{ratio:} &\quad e^{\beta_1(x_1+1-x_1)} = e^{\beta_1} \end{aligned}$$

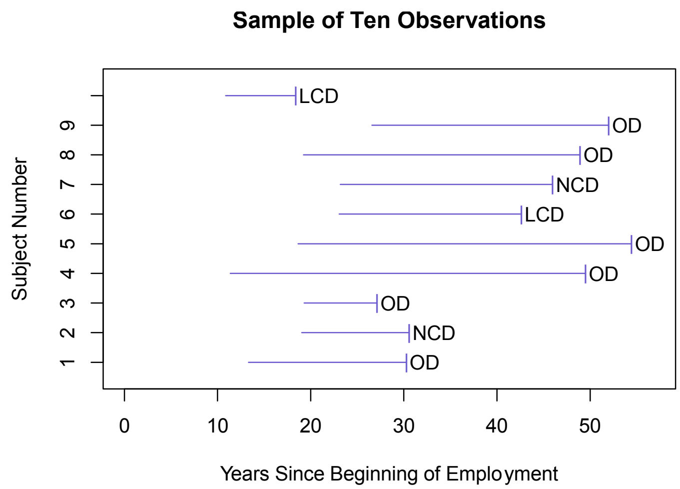
COX REGRESSION MODEL

- $e^{\beta 1}$ is the RR associated with a one-unit difference of $x_{1,}$ holding other x's and t constant.
- Some functional form is required for how the hazard function at each t depends on $x_2 \ldots x_k$.
- No functional form is required for how the hazard at each $x_2 \ldots x_k$ depends on t, since $\lambda_0(t)$ can be any function.
- The time scale for t is the variable that is adjusted for the most finely/thoroughly.

WELSH NICKEL REFINERS



WELSH NICKEL REFINERS



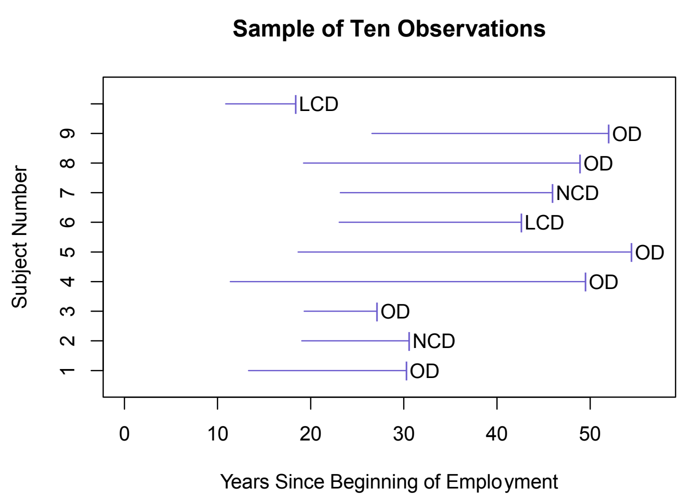
OBSERVATION STARTING LATE

- Should not include subjects in risk sets before they are under observation:
  - Other subjects "just like" them who died before their entry time are not observed
  - Falsely inflates the numbers at risk in early risk sets
  - Biases cause-specific hazard estimation
  - Can bias Cox model estimation

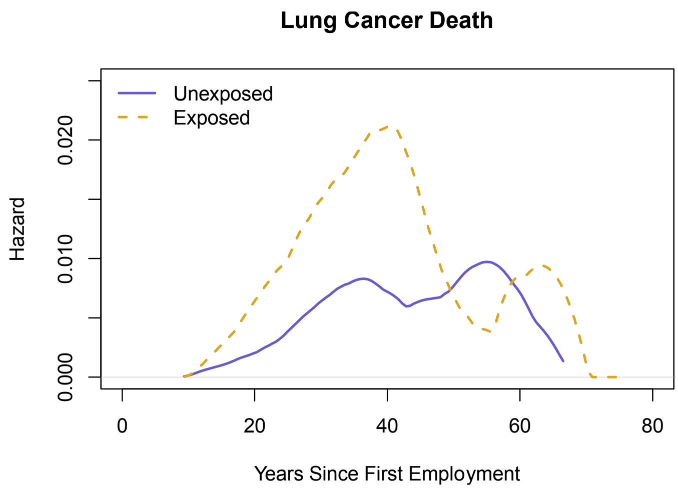
OBSERVATION STARTING LATE

- Solution: "Left enter" subjects at time when active follow-up starts
  - Subjects only contribute to risk sets where their event could have been observed
  - They are only in the denominator if we could have seen them in the numerator

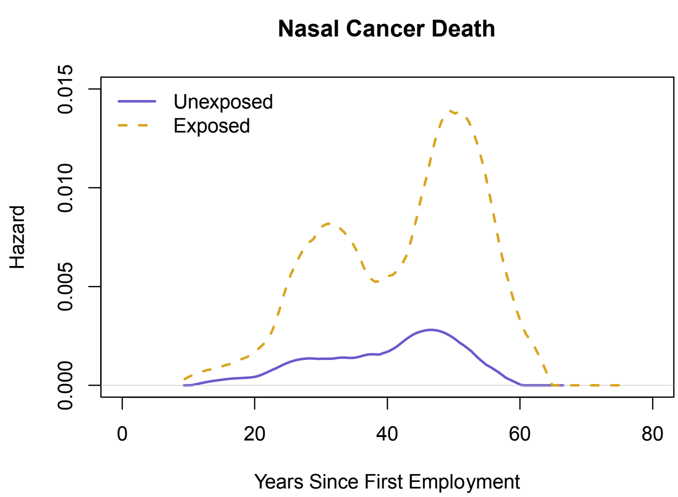
WELSH NICKEL REFINERS



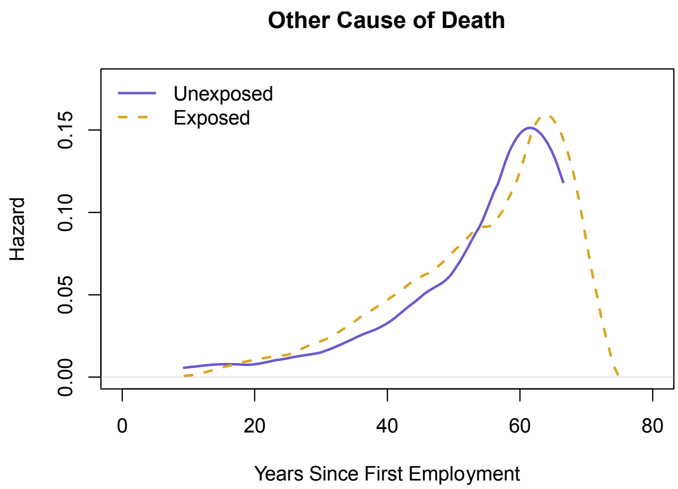
LUNG CANCER



NASAL CANCER



OTHER CAUSES



LUNG CANCER

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
exposedTRUE	0.8000334	2.225615	0.1860041	4.301159	1.7e-05

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
exp0.5 - 4.0	0.6111674	1.842581	0.2123734	2.877796	0.0040046
exp4.5 - 8.0	1.0952795	2.990018	0.2838639	3.858467	0.0001141
exp8.5-12.0	1.2880174	3.625591	0.3739070	3.444754	0.0005716
exp12.5+	1.4327121	4.190048	0.4791166	2.990321	0.0027868
afe20-27.5	0.7604881	2.139320	0.3081636	2.467806	0.0135944
afe27.5 - 35	0.8670846	2.379962	0.3281099	2.642665	0.0082256
afe35+	0.7982183	2.221579	0.4224336	1.889571	0.0588154
yfe1910-1914	0.4358460	1.546271	0.2724801	1.599552	0.1096981
yfe1915-1919	0.1753274	1.191636	0.3775109	0.464430	0.6423397
yfe1920-1925	0.6547157	1.924595	0.2991155	2.188839	0.0286086

NASAL CANCER

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
exposedTRUE	1.540408	4.666495	0.3503185	4.397165	1.1e-05

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
exp0.5 - 4.0	0.8958359	2.449382	0.4044464	2.2149680	0.0267623
exp4.5 - 8.0	1.1991717	3.317368	0.4727052	2.5368277	0.0111862
exp8.5-12.0	2.3214816	10.190761	0.5173928	4.4868842	0.0000072
exp12.5+	2.8655920	17.559445	0.5727364	5.0033346	0.0000006
afe20-27.5	1.4721869	4.358757	0.7527320	1.9557917	0.0504897
afe27.5 - 35	2.1770312	8.820082	0.7601145	2.8640834	0.0041822
afe35+	3.6025888	36.693104	0.7886401	4.5681026	0.0000049
yfe1910-1914	1.0373701	2.821786	0.3798834	2.7307593	0.0063189
yfe1915-1919	1.1291520	3.093033	0.5130845	2.2007137	0.0277563
yfe1920-1925	0.0166965	1.016837	0.5257787	0.0317558	0.9746668

OTHER CAUSE OF DEATH

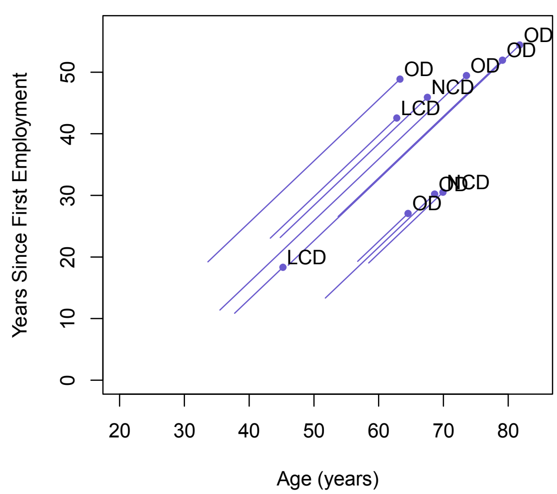
	coef	exp(coef)	se(coef)	z	Pr(>|z|)
exposedTRUE	0.2164895	1.24171	0.0966131	2.240788	0.0250398

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
exp0.5 - 4.0	0.1685250	1.183558	0.1106070	1.5236376	0.1275993
exp4.5 - 8.0	0.2360561	1.266245	0.1602288	1.4732445	0.1406851
exp8.5-12.0	0.0585201	1.060266	0.2564181	0.2282213	0.8194742
exp12.5+	0.0245456	1.024849	0.3964995	0.0619059	0.9506378
afe20-27.5	0.5704774	1.769111	0.1545876	3.6903186	0.0002240
afe27.5 - 35	1.1656136	3.207891	0.1665088	7.0003136	0.0000000
afe35+	2.0835886	8.033245	0.1957375	10.6448086	0.0000000
yfe1910-1914	0.2087081	1.232085	0.1540413	1.3548842	0.1754544
yfe1915-1919	0.2329453	1.262312	0.1788233	1.3026563	0.1926921
yfe1920-1925	0.1024386	1.107869	0.1529133	0.6699127	0.5029135

CHOOSING A TIME SCALE

- What time scale makes the most sense for the Welsh Nickel Refiners study?

TWO TIME SCALES



CHOOSING A TIME SCALE

- Cardiovascular Health Study
  - NHLBI cohort of older Americans (65+)
  - Many baseline demographic and health measures.
  - Follow-up for more than 20 years for a large number of health conditions.
- What is the best time scale: age or time since baseline?

TIME INTERACTIONS

- So far, all our Cox models have assumed that the hazard ratio is constant over time
- It's possible to incorporate interaction terms with functions of time to allow the HR to depend on time.

TIME INTERACTIONS

One way for the hazard ratio to depend on time: interaction with a function of time t.

$$\lambda(t) = \lambda_0(t)e^{\beta_1 x + \beta_2 x f(t)}$$

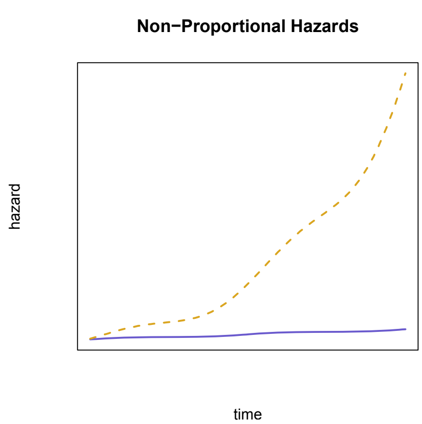
Here the hazard ratio depends on time through the interaction term:

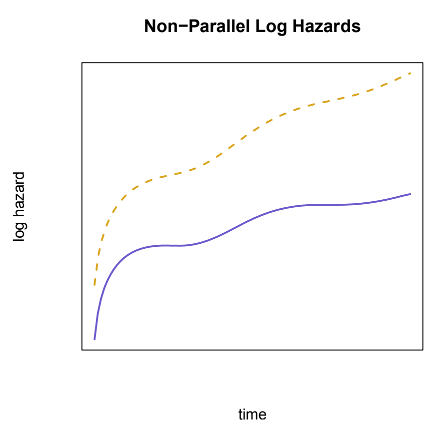
$$\begin{aligned}
\lambda(t|x+1) &= \lambda_0(t)e^{\beta_1(x+1)+\beta_2(x+1)\cdot f(t)} \\
\lambda(t|x) &= \lambda_0(t)e^{\beta_1 x+\beta_2 x\cdot f(t)} \\
\text{hazard ratio} &= e^{\beta_1(x+1)+\beta_2(x+1)\cdot f(t)-\beta_1 x-\beta_2 x\cdot f(t)} = e^{\beta_1+\beta_2 f(t)}
\end{aligned}$$

Requires a hypothesized functional form for the dependence of the hazard ratio at time t on t.

Commonly-used functions of t are $f(t) = t$ and $f(t) = \log(t)$.

TIME INTERACTIONS



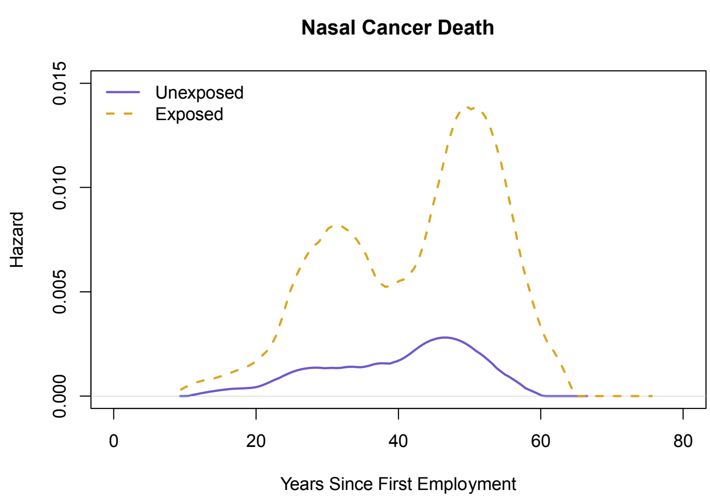


NASAL CANCER TIME INTERACTION

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
exposedTRUE	1.540408	4.666495	0.3503185	4.397165	1.1e-05

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
exposedTRUE	1.0613334	2.890222	5.161871	0.2056102	0.8370954
tt(exposed)	0.1290554	1.137753	1.388229	0.0929641	0.9259321

NASAL CANCER



ESTIMATING THE HR AS A FUNCTION OF TIME

- In exploratory analyses, may be of interest to estimate how the hazard ratio varies over time
- Estimate based on ratio of kernel-smoothed hazard estimates can be very variable
- Better choice is based on smoothed Schoenfeld residuals
- Can be thought of as an estimate of a time-dependent coefficient of a fixed variable

ESTIMATING THE HR AS A FUNCTION OF TIME

Another way for the hazard ratio to depend on time: time-dependent coefficients.

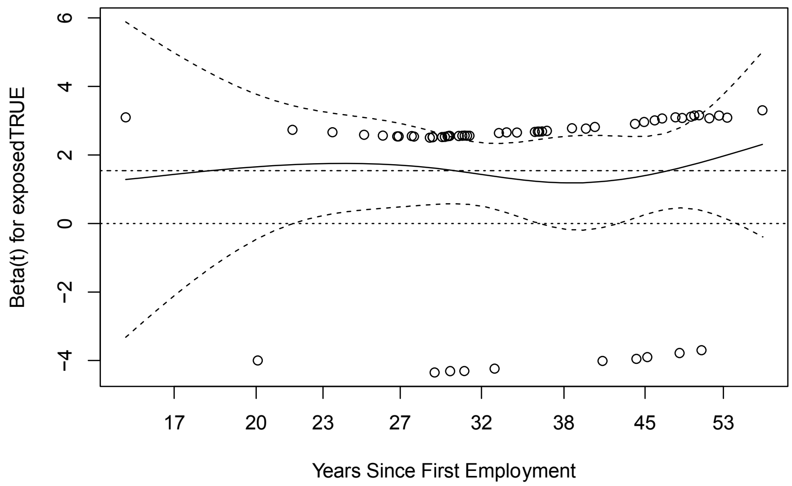
$$\lambda(t) = \lambda_0(t)e^{\beta(t)x}$$

Here the hazard ratio depends on time through the time-dependent coefficient $\beta(t)$

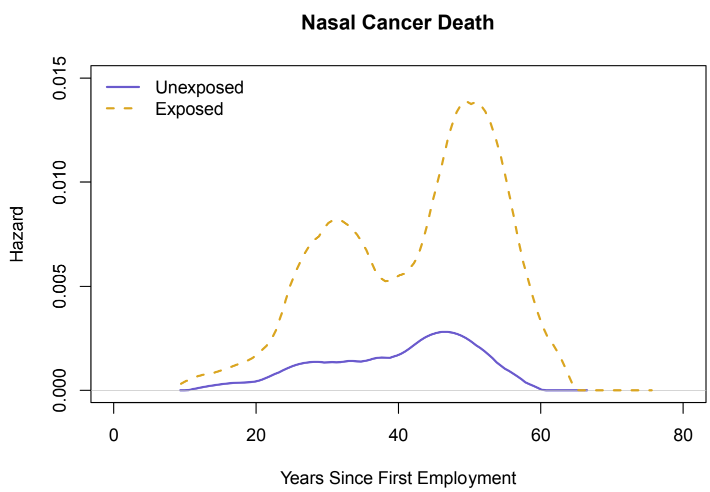
$$\begin{aligned} \lambda(t|x+1) &= \lambda_0(t)e^{\beta(t)(x+1)} \\ \lambda(t|x) &= \lambda_0(t)e^{\beta(t)(x)} \\ \text{hazard ratio} &= e^{\beta(t)(x+1)-\beta(t)x} = e^{\beta(t)} \end{aligned}$$

Estimated hazard ratio can be an arbitrary function of time $e^{\beta(t)}$.

NASAL HR ESTIMATE



NASAL CANCER



EXAMPLE

- Real et al. 2016, “Survival Predictors in Liver Transplantation: Time-Varying Effect of Red blood Cell Transfusion”, Transplantation Proceedings, 48, 3303.
- 543 consecutive patients, 2006-2014, retrospectively
- Preoperative
  - Age, sex, Model for End-Stage Liver Disease score, primary diagnosis, cold ischemia time, international normalized ratio, serum albumin, hemoglobin levels
- Intraoperative
  - Norepinephrine, blood loss, red blood cell transfusions surgical time

RESULTS

- Only significant independent predictors:
- Red blood cell transfusion, HR=1.16 (1.04-1.29)
- Sex, HR=1.71 (1.10-2.65)
- Non-proportionality
  - “multivariate Cox regression model was subsequently upgraded by adding a time-verying interaction between red blood cell transfusion and time since liver transplantation”

RESULTS

Table 3. Multivariate Cox Regression (Time-Varying Interaction With RBC Transfusion Not Included)

Variable	HR	95% CI	*P* Value
Age (y)	1.02	1.00–1.04	.061
Female sex*	1.71	1.10–2.65	.016
Intraoperative RBC (units)	1.16	1.04–1.29	.005
Intraoperative blood loss (L)	0.90	0.79–1.03	.135
Intraoperative norepinephrine (mg)	1.02	0.97–1.07	.508
Surgical time (h)	1.19	0.95–1.48	.125

Abbreviations as in Tables 1 and 2.
*$P < .05$.

RESULTS

Table 4. Multivariate Cox Regression Including the Time-Varying Interaction With RBC Transfusion

Variable	HR	95% CI	*P* Value
Age (y)	1.02	1.00–1.04	.077
Female sex	1.66	1.07–2.56	.024
Intraoperative RBC (units)	1.25	1.12–1.40	.000
Intraoperative blood loss (L)	0.91	0.80–1.03	.147
Intraoperative norepinephrine (mg)	1.01	0.96–1.06	.803
Surgical time (h)	1.20	0.96–1.49	.105
Time-varying interaction with intraoperative RBC transfusion	0.98	0.97–0.99	.001

Abbreviations: as in Tables 1 and 2.

RESULTS

Table 5. Time-Varying Effect of RBC Transfusion on Patient Survival

Time Since LT	HR	95% CI	*P* Value
3 mo	1.14	1.020–1.257	.015
6 mo	1.12	1.003–1.240	.033
1 y	1.11	0.986–1.225	.070
2 y	1.09	0.968–1.210	.132
3 y	1.08	0.958–1.202	.183

Abbreviations: LT, liver transplantation; others as in Tables 1 and 2.