Choosing time scale in a competing risk setting when using Flexible Parametric Models

SKOURLIS NIKOLAOS

Research question

- Competing risks setting (Cause I :Death due to cancer, Cause II: Death due to other causes)
- Choice of timescale when modeling each event

Death due to cancer : Time since diagnosis

Death due to other causes: Attained age

- Use of the ”wrong” timescale/time since diagnosis for cause II :

Does the modeling of age at diagnosis play a role in the resulting bias?

Natural modeling approach for timescale:
Cause 2: Attained age : $h_2(a|x)$, $a = a_0 + t$

”Wrong” timescale approaches

- Cause 2: Time since diagnosis: $h_1(t|a_0, x)$, a_0 linear
- Cause 2: time since diagnosis: $h_1(t|a_0, x)$, a_0 splines
- Cause 2: time since diagnosis: $h_1(t|a_0, x)$, a_0 splines plus time- age interaction terms

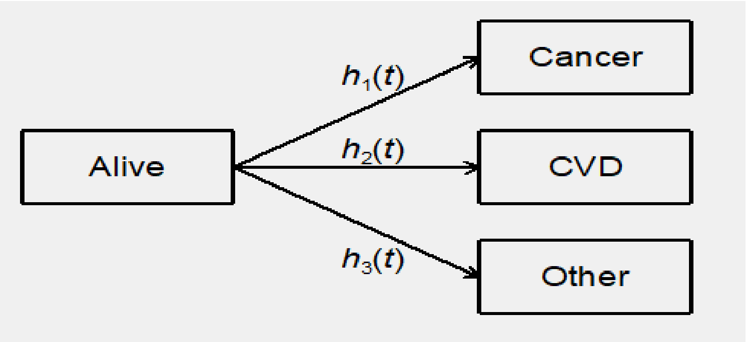
Competing risk setting

Survival analysis that aims to correctly estimate the marginal probability of an event in the presence of competing events

Each competing event is an absorbing state

Estimation of probability of each competing event taking into account the risk of all potential events (estimation of CIFs)

<u>Cummulative Incidence Function (CIF):</u> marginal probability of a certain event as a function of its cause-specific probability and overall survival probability



$$S_k(t) = exp(-(\int_0^t h_k(u)\, du))$$

$$CIF_k(t) = \int_0^t S_1(u)\, S_2(u)\, h_k(u)\, du \quad , \quad k = 1,2$$

Competing risk setting
Modeling on attained age for death due to other causes

There is no strict restriction as to which time-scale can be used for modelling each event.

In a setting with 2 potential events (Death due to cancer, Death due to other causes) we have alternatives:

- Both events modeled under time since diagnosis timescale (Most frequent approach)

Time since diagnosis (t) ➔ Death due to cancer: $h_1(t|a_0, X)$

Time since diagnosis (t) ➔ Death due to other : $h_2(t|a_0, X)$

- Death due to cancer modeled with time since diagnosis timescale and Death due to other causes with attained age

Time since diagnosis (t) ➔ Death due to cancer: $h_1(t|a_0, X)$

Attained age ($a = a_0$+t) ➔ Death due to other : $h_2(a|X) = h_2(a_0\text{+t}|X)$

Competing risks and Relative survival framework

All cause hazard can be partitioned to hazard of dying from cancer and hazard of dying due to other causes:

$$h_{all\ cause}(t) = h_{other\ causes}(t) + h_{cancer}(t) \quad (1)$$

1. In a relative survival the type 1 takes the form:

$$h_{all\ cause}(t) = h^*(t) + \lambda_{cancer}(t)$$

- the expected mortality $h^*(t)$ and the all cause mortality $h_{all\ cause}$ are considered known, derived directly by the lifetables of popmort files
- Only the excess mortality $\lambda_{cancer}(t)$ is modelled
- Cause of death information is avoided

2. In a competing risks setting, type 1 keeps the form:

$$h_{all\ cause}(t) = h_{other\ causes}(t) + h_{cancer}(t)$$

- Both the hazard of dying from cancer $h_{cancer}(t)$ and hazard of dying due to other causes $h_{other\ causes}(t)$ are modeled
- Cause of death information is used for modelling of both causes

Competing risks and Relative survival framework
Crude probabilities from both settings

- Crude probabilities is the "natural" end product of a competing risk analysis ($CIFs$)
- Net probabilities is the "natural" end product of a relative survival analysis ($RS, Net\ P_{cancer}$)
- Crude estimates can also be derived from a RS framework: crude probabilities due to cancer and due to other causes can be estimated from life tables Cronin and Feuer (2000) or from excess mortality models Lambert et al. (2010) evading the death certificates issues (*Paul Dickman, Enzo Coviello, 2015: Estimating and modeling relative survival, Section 4.8)*
- *Competing risks (CIF for death due to cancer) with attained age as timescale for death due to other causes:*

$$CIF1(t|A=a_0,X) = \int_0^t S_1(u|a_0,X) * S_2(u|a_0,X) * h_1(u|a_0,X) \xrightarrow{\text{attained age for cause 2}} \int_0^t S_1(u|a_0,X) * \frac{S_2(a_0+u|X)}{S_2(a_0|X)} * h_1(u|a_0)$$

- *Relative survival setting (crude probability of death due to cancer):*

$$CIF1(t|A=a_0,X) = \int_0^t RS(u|a_0,X) * S^*(t|a_0,X) * \lambda(u|a_0,X) \longrightarrow \int_0^t RS(u|a_0,X) * \frac{S^*(a_0+u|X)}{S^*(a_0|X)} * \lambda(u|a_0,X)$$

Background

❖Korn et al (1997) suggested 2 conditions under which attained age and time since diagnosis approaches should give same estimates

- Baseline hazard is an exponential function of time
- Even if not, the effect estimates should be very close if covariate X is independent of baseline age a0

❖Benichou et al, 2004

- If X a0, then no bias due to confounding but still potential bias towards null if model misspecification of a0
- If baseline hazard not exponential→ upwards confounding bias of age ta baseline but quite small

❖ Chalise et al 2012 notes:

- Baseline hazard follows gompertz → attained age vs time since diagnosis- linearly adjusted for baseline age approach should give the same results
- When the chronological age is the correct timescale, the time on study time-scale model is reasonably close to the attained age time-scale model.

$$\begin{aligned}\lambda_A(a|x) &= \lambda_{0A}(a)e^{\beta x}\\ &= ce^{\gamma a}e^{\beta x}\\ &= ce^{\gamma(a_0+t)}e^{\beta x}\\ &= ce^{\gamma t}e^{\beta x+\gamma a_0},\end{aligned}$$

* The bibliography is based on regular survival analysis where one event is studied and the effect of interest is the beta coefficient

Scenarios overview- Compared approaches

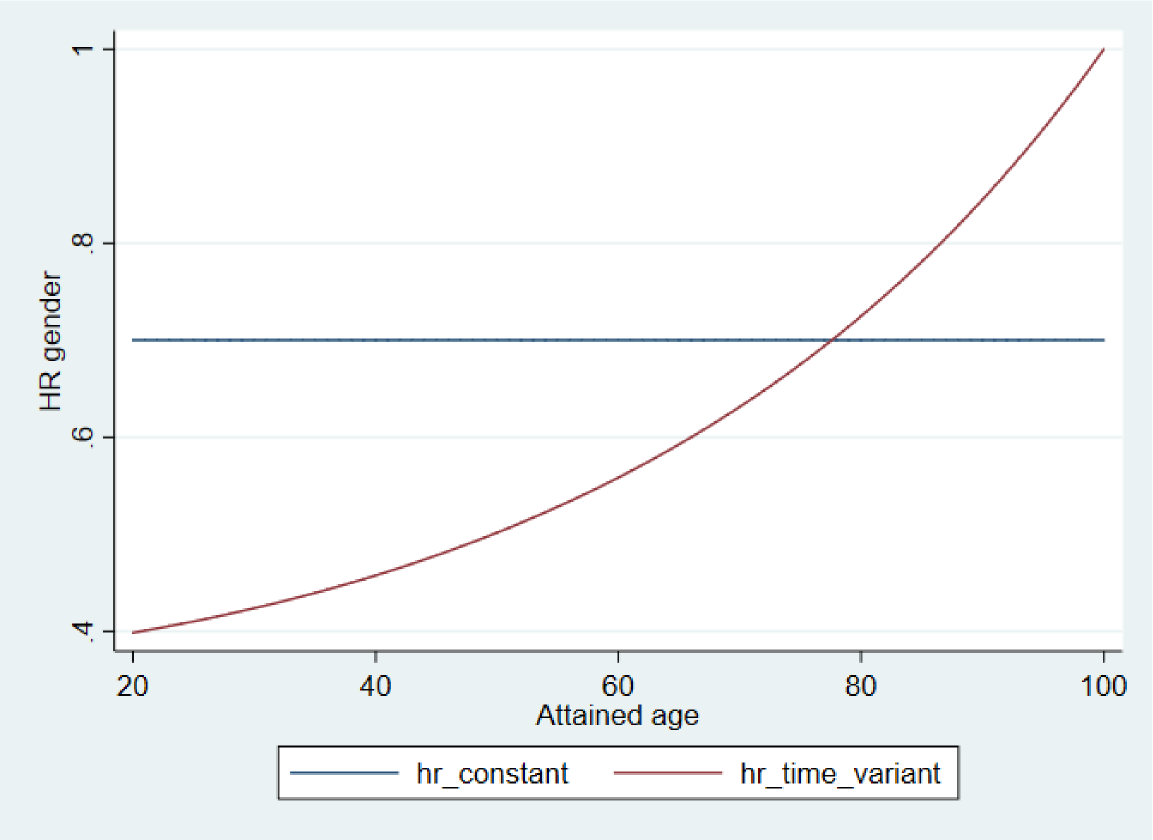
Scenarios For cause I	Baseline hazard On time since diagnosis	Age-Gender dependence	HR of gender for cause 2
1 , 2	Mixture of weibulls Quadratic effect of age	Yes/No	Constant HR= 0.95

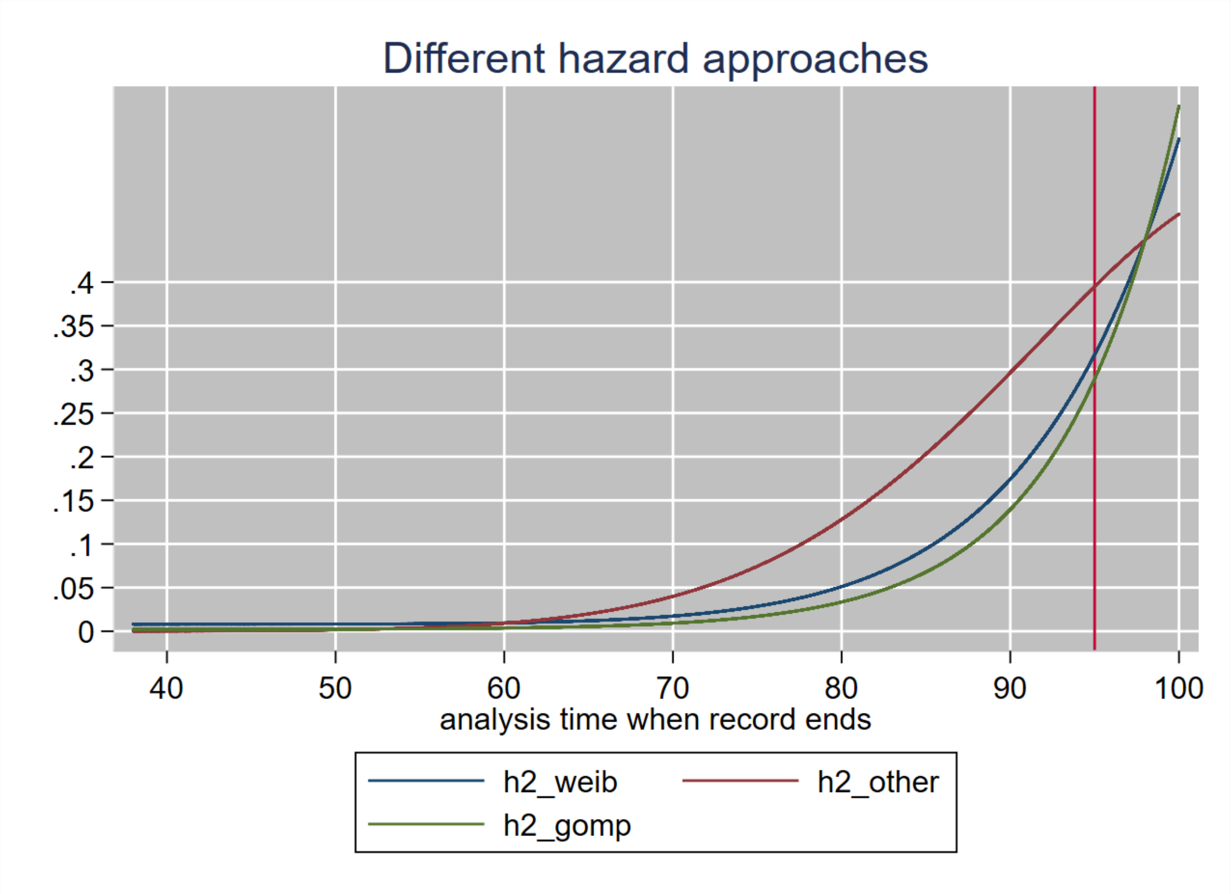
Scenarios For cause II	Baseline hazard On attained age	Age-Gender dependence	HR of gender for cause 2
1	Weibull	Yes	Constant HR= 0.7
2	Weibull	Yes	Time varying HR: 0.4 at 20 to 1 at 100 of attained age
3	Weibull	No	Constant HR= 0.7
4	Weibull	No	Time varying HR: 0.4 at 20 to 1 at 100 of attained age
5	Other hazard shape	Yes	Constant HR= 0.7
6	Other hazard shape	Yes	Time varying HR: 0.4 at 20 to 1 at 100 of attained age
7	Other hazard shape	No	Constant HR= 0.7
8	Other hazard shape	No	Time varying HR: 0.4 at 20 to 1 at 100 of attained age
9	Gompertz	Yes	Constant HR= 0.7
10	Gompertz	Yes	Time varying HR: 0.4 at 20 to 1 at 100 of attained age
11	Gompertz	No	Constant HR= 0.7
12	Gompertz	No	Time varying HR: 0.4 at 20 to 1 at 100 of attained age

Cause I	Timescale	Modelling of age at diagnosis
Only approach	Time since diagnosis	4 spline terms

Cause II	Timescale	Modelling of age at diagnosis
1st approach (“correct”)	Attained age	-
2nd approach	Time since diagnosis	Linear term
3rd approach	Time since diagnosis	4 spline terms
4th approach	Time since diagnosis	4 spline terms for the main effect+ 3 spline terms for the age at diagnosis-time since diagnosis interaction

Effect of gender, Baseline hazards, Age-gender dependence scenarios





- Age at diagnosis-Gender independence: Age at diagnosis~N (65,15)
- Age at diagnosis-Gender dependence: Age at diagnosis~N (63,15) for males, Age at diagnosis~N (67,15) for females

Simulation results overview: Use of RShiny interactive graphs

For each scenario, for t=1,2,3,4,5,6,7,8,9,10 years after diagnosis, for males and females over ages at diagnosis 50,60,70,80,90

- Bias in CIF1 and CIF2
- Monte Carlo error in estimations
- % Coverage of true CIF values
- Relative efficiency compared to attained age approach
- Estimated HR of gender
- Estimated CIF differences and ratios (males vs females) from each approach and comparison with truth
- Convergence of each model

Results-Discussion

1. The linear age term approach for cause 2 leads to heavily biased results both for CIF2 for most scenarios as expected.
2. All approaches are unbiased for CIF1 (at α0=90 the linear approach heavily biased)
3. In non PH scenarios, for α0=60 and 90 and hazard ”Other”, the bias in CIF2 for females under the single timescale approaches (splines, spline+interaction) is noticeably bigger compared to that of the ”standard” approach
4. For extreme ages (α0=90), the bias in CIF1 appears to smaller in the single timescale approaches (splines, splines+interaction)

- Even if we model death due to other causes with the ”wrong” underlying timescale, we will not necessarily get bigger bias compared to modeling using the correct timescale, provided we include the effect of age at diagnosis in the appropriate way
- We argue that using the attained age as underlying timescale when this is the ”natural” choice, will result to an unbiased- simple model, less prone to misspecification.