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

SKOURLIS NIKOLAOS

Contents

- Competing risks setting with use of CSH models
- On the choice of timescale when modeling each event
- Competing risks with attained age for cause 2 and Relative survival framework- brief comparison
- On the impact of choosing the ”wrong” timescale- When it should not matter
- Use of the ”wrong” timescale: Does the choice of modeling approach play a role in the resulting bias?
- Building our simulation- Scenarios-Modeling approaches
- Results and Discussion

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

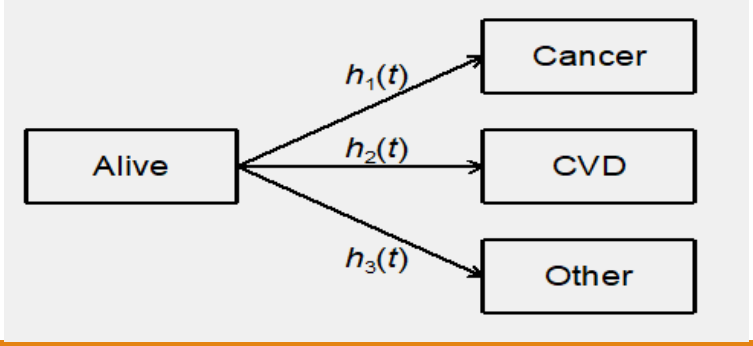
Approaches for modelling the hazards:

a) Subhazards approach

b) Cause specific hazard model for each event/state (CSH)

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

Cummulative Incidence Function (CIF): 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 : $h_2(a|X) = h_2(a_0$+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 ($RS, Net\ P_{cancer}$) $RS(t) = \frac{S(t)}{S^*(t)}$ $Net\ P_{cancer} = 1 - RS(t)$
- 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:*

$$CIF1(t|A = a_0, X) = \int_0^t S_1(u|a_0, X) * S_2(u|a_0, X) * h_1(u|a_0, X)$$

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):*

$$CIF1(t|A = a_0, X) = \int_0^t RS(u|a_0, X) * S^*(t|a_0, X) * \lambda(u|a_0, X)$$

$$\int_0^t RS(u|a_0, X) * \frac{s^*(a_0 + u|X)}{s^*(a_0|X)} * \lambda(u|a_0, X)$$

Timescale selection for death due to other causes
When it should not matter

- Korn et al. (1997) suggested two conditions where attained age and time since diagnosis approach should give same estimates

the baseline hazard is approximately exponential

Even if baseline hazard is not exponential, the effect estimates would be very close if the covariate z is independent of the baseline age a0 ***(z*** $\amalg a_0$ ***)***.

- Benichou et al 2004

If ***z*** $\amalg a_0$ ➔ no bias due to confounding but still potential bias towards null if model misspecification of a_0 in the model

If baseline hazard not exponential ➔ upwards confounding bias of age but quite small

*the bibliography is based on regular survival analysis where one event is studied, and the bias is reffering to the estimation of the effect of gender in PH Cox models

Timescale selection for death due to other causes
When it should not matter

- Pencina et al. (2007) notes:

if all ages at baseline were the same, the models would all be equivalent. The distribution of ages at baseline plays a central role in determining when the time scales will produce equivalent results.

- 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.

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

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

Use of the ”wrong” timescale: Does the choice of modeling approach play a role?
Building our simulation

- Simulate random datasets with age at diagnosis and gender variables close to the population distribution
- Simulate survival times from :

  Mixture of Weibulls ➔ Death due to cancer (Time since diagnosis)

  Weibull/ Gompertz- Makeham/ Other ➔ Death due to other causes (Attained age)

  Keep the event with the minimum survival time
- Derive the true values of CIF1(t) and CIF2(t)

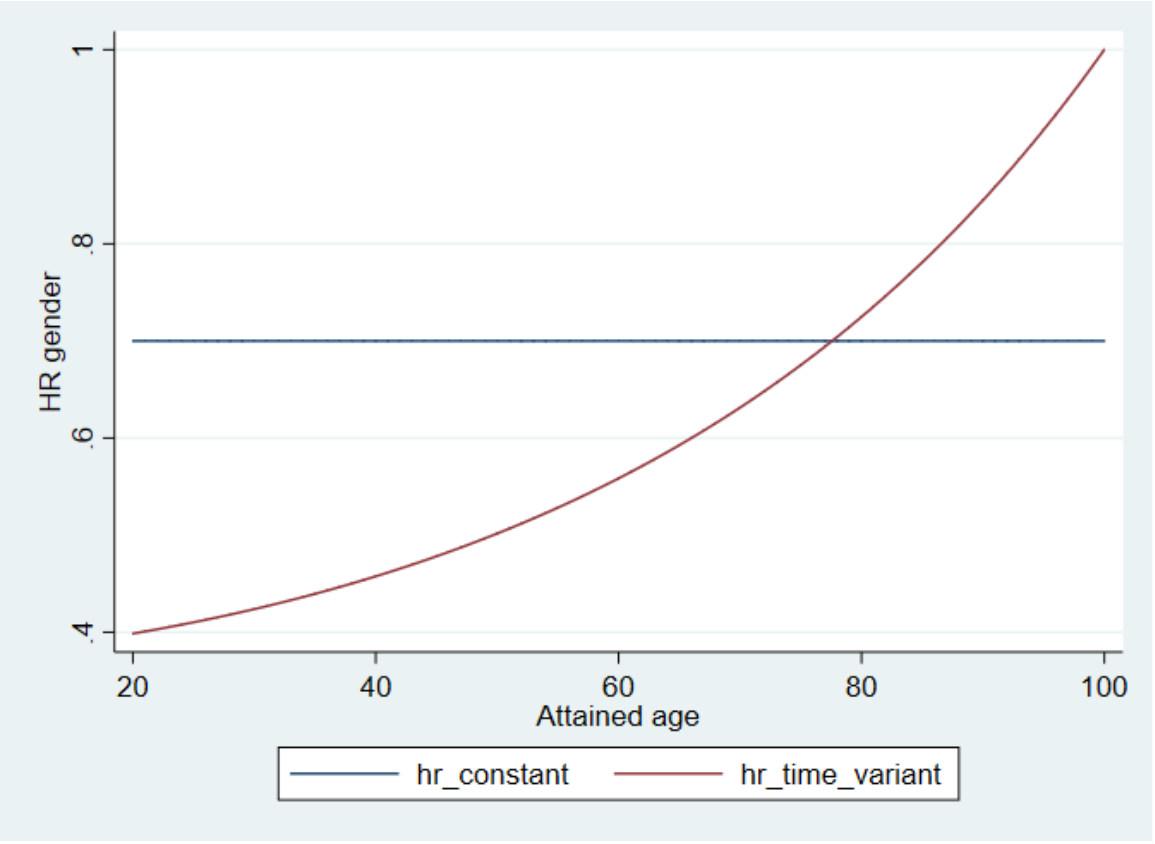
Scenarios

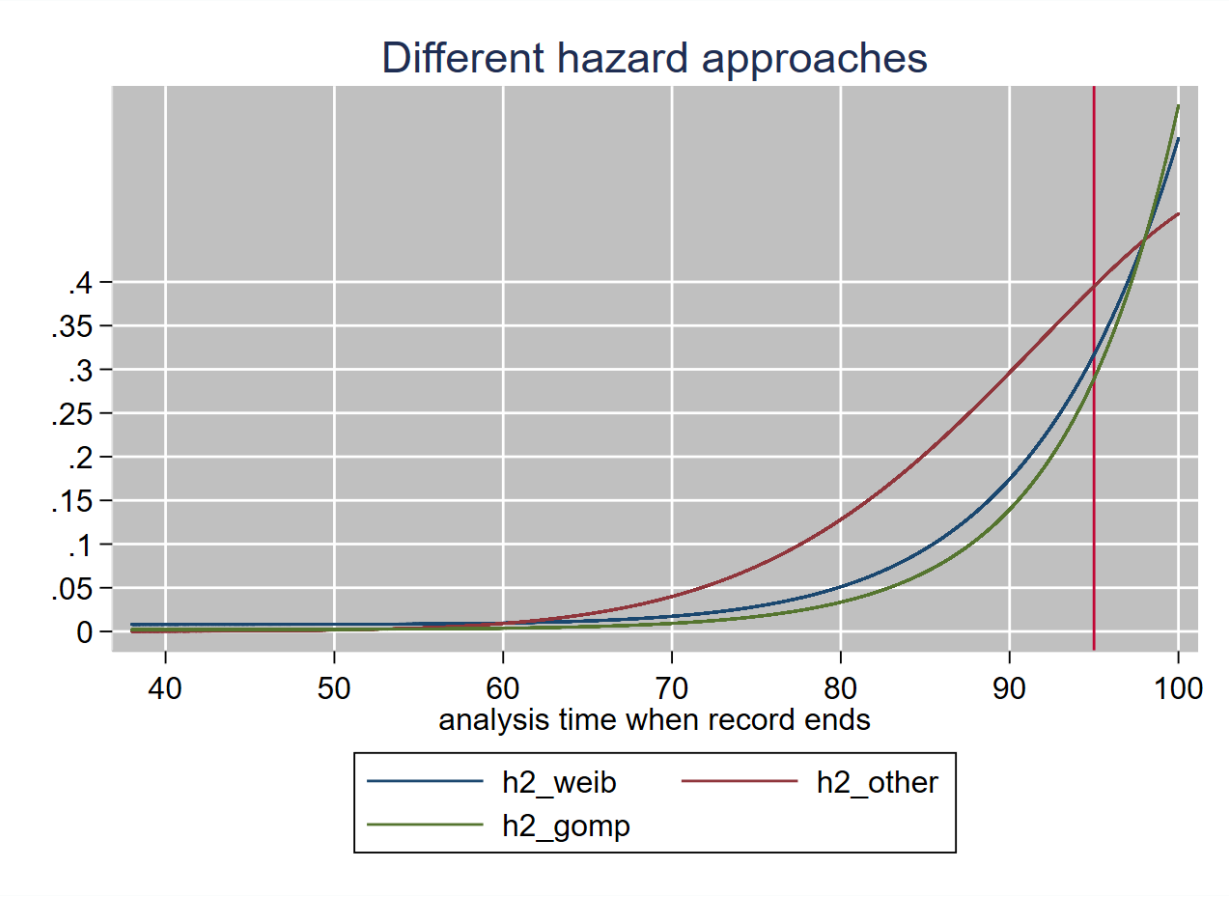
1. Different effects of gender for cause II (Case I: constant HR=0.7, Case 2: Quardratic shape of HR from 0.4 at 20 years to 1 at 100 years)
2. Different distributional assumptions for the baseline hazard (Weibull/Gompertz/Other)
3. Independence or dependence of age at diagnosis and gender. (Different age distribution mean across genders if dependence of age at diagnosis with gender)

Total of 12 scenarios

- Only 1 scenario for cause I : Quadratic effect of age at diagnosis (Beta1=0.037, Beta2=0.0006) and HR_{gender}=0.95

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

Scenarios overview

Scenarios	Baseline hazard	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

Estimating the CIFs

Models

Flexible parametric models:

- Time s. diag for cause I- Attained age for cause II, only gender as covariate (”Correct” approach)
- Time s. diag for both causes, gender as covariate, age at diagnosis as covariate (linear, spline terms, splines+interaction)

Models	Age at diagnosis	Gender	Comments
Cause I -Model	Spline terms	PH of gender	Common cause I modeling in all approaches
Cause II- time s. diagnosis	Linear term	Allow/ Don’t allow for time varying effects	More naïve model expecting bias
Cause II- time s. diagnosis	Spline terms	Allow/ Don’t allow for time varying effects	
Cause II- time s. diagnosis	Spline terms plus age-time interaction terms	Allow/ Don’t allow for time varying effects	
Cause II – attained age	-	Allow/ Don’t allow for time varying effects	“Correct” model DGM based on this model

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

Simulation results overview: Use of RShiny interactive graphs

Open Rshiny App

Results

1. The linear age term approach for cause 2 leads to heavily biased results both for CIF2 for most scenarios as expected.
2. Under all baseline hazard: All approaches are unbiased for CIF1 (at α_0=90 the linear approach heavily biased)
3. In non PH scenarios, for α_0=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 (α_0=90), the bias in CIF1 appears to smaller in the single timescale approaches (splines, splines+interaction)
5. Age at diagnosis – Gender Dependence does not seem to have any systematic effect on bias.

In some cases dependence results to a reduction in bias (possible explanation: more events to the extreme ages due to an increased total variance in the total age at diagnosis distribution)

Discussion

- 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:

1. Age at diagnosis spline terms (age at diagnosis-time interaction spline terms may contribute)
2. time varying effects for the effect of covariates,

In some cases, we may even have more stable results for the extreme ages (>90) for CIF1 under the single timescale approaches

In non PH scenarios, the common time scale approach can have small biases in CIF2 estimation even if we have fully model age at diagnosis and have tvc terms for gender.

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.

Thanks!!!