Modèle de regression pour la probabilité conditionelle d'un événement concurrent.

Arthur Allignol, Jason P. Fine, Jun Yan,
Aurélien Latouche (UVSQ, EA2506)

Contexte

- Essai Clinique : 2 groupes (E/C)
- Critère de Jugement Censuré (Rechute)
- On observe aussi l'indicatrice du type d'evt ε
- Evts concurrents/compétitif: 2 evts exclusifs ($\varepsilon = 1, 2$)
- Evts censurés ($\varepsilon = 0$)
- Objets familliés
  - Risque cause spécifique
  - Incidence cumulée et *variantes*

Intérêt : **Effet** de (co)variables sur ces quantités (critère de jugement principal)

Fonctions de risques *Hazards functions*

- Cause–spécifique (CSH)

$$\lambda_{1.}(t) = dF_{1.}(t)/S(t)$$

- *Subdistribution* (SH)

$$\alpha_{1.}(t) = dF_{1.}(t)/(1 - F_{1.}(t))$$

où

- $F_{1.}(t) = \Pr(T \leq t, \epsilon = 1)$ Incidence cumulée
- $S_{.}(t) = 1 - (F_{1.}(t) + F_{2.}(t))$ *event free survival*

Rq: $\lambda_1 > \alpha_1$ et F_1 non linéaire en λ_i

Modèles de régression à risques proportionnels

Relation entre une variable réponse (ici un délai d'évènement, T) et des covariables explicatives

2 quantités fondamentales :

1. La fonction de risque cause-spécifique (CSH)
2. La fonction d'incidence cumulée (CIF)

Modèles semi-paramétriques à risques proportionnels :

$$\underbrace{\lambda(t;Z)}_{\text{Fct de risque}} = \underbrace{\lambda_0(t)}_{\text{Risque de base}} \exp(\beta Z)$$

$$\frac{\lambda(t;Z=1)}{\lambda(t;Z=0)} = \exp(\beta), \text{ rapport des risques } (\textit{hazard ratio})$$

Risques proportionels

Modèles à risques proportionnels pour ces 2 fonctions :

- Cox : $\lambda_i(t; Z) = \lambda_{i0}(t) \exp(\beta_i Z)$
- Fine–Gray : $\alpha_i(t; Z) = \alpha_{i0}(t) \exp(\gamma_i Z)$

Implémenté `survival`, `cmprsk`

Interpretation

- Une variable peut n'avoir aucun effet sur les CSH (λ_1, λ_2)
- Et avoir un effet sur F_1
- Résultats discordants
- A risk factor may decrease all cause–specific hazard, but lead to an increase in terms of the CIF is somewhat counterintuitive.
- Interprétation clinique ?
- Effet physiologique sur l'evt d'interêt (ou sur l'evt concurrent)
- Modèles mal spécifié (Latouche, 2007, Stat Med)

Hypothèse de proportionnalité

Soit h une fonction de risque (*hazard*) , la proportionnalité s'exprime :

$$h_{1E}(t) = h_{1C}(t) \text{ HR}$$

où HR est le rapport des risques (*hazard ratio*)

$$H_0 : h_{1E}(t) = h_{1C}(t) \iff \text{HR} = 1.$$

Test(s) du log–rank

H_0: les probabilités instantanées de décès sont les mêmes entre les deux groupes

- Analyse de survie : équivalence entre risque et survie
- Test fondé sur les différences pondérées integrées entre
  - Risque cumulé (CSH) estimé dans le bras de traitement
  - Risque cumulé estimé en réunissant les 2 bras de traitement
- Tarone–Ware (1977), Fleming Harrington (1981)
- Ces tests sont implémentés donc utilisés
- Optimaux pour des *risques* (CSH) proportionnels

Comparaison d'incidences cumulées

$H_0 : F_{1E} = F_{1C}$

De la forme

$$Z_\cdot = \int_0^\tau W_\cdot(u) \left\{ d\hat{\Gamma}_{1\cdot}(u) - d\hat{\Gamma}_1^0(u) \right\},$$

- τ la durée totale de l'essai
- $\hat{\Gamma}_{1\cdot}(t)$ Risque (SH) cumulée estimé dans le bras de traitement "$\cdot$"
- $\hat{\Gamma}_1^0(t)$ Risque (SH) cumulée estimé en réunissant les 2 bras de traitement

Test implémenté dans `R` et optimal pour des SH proportionnels [a]

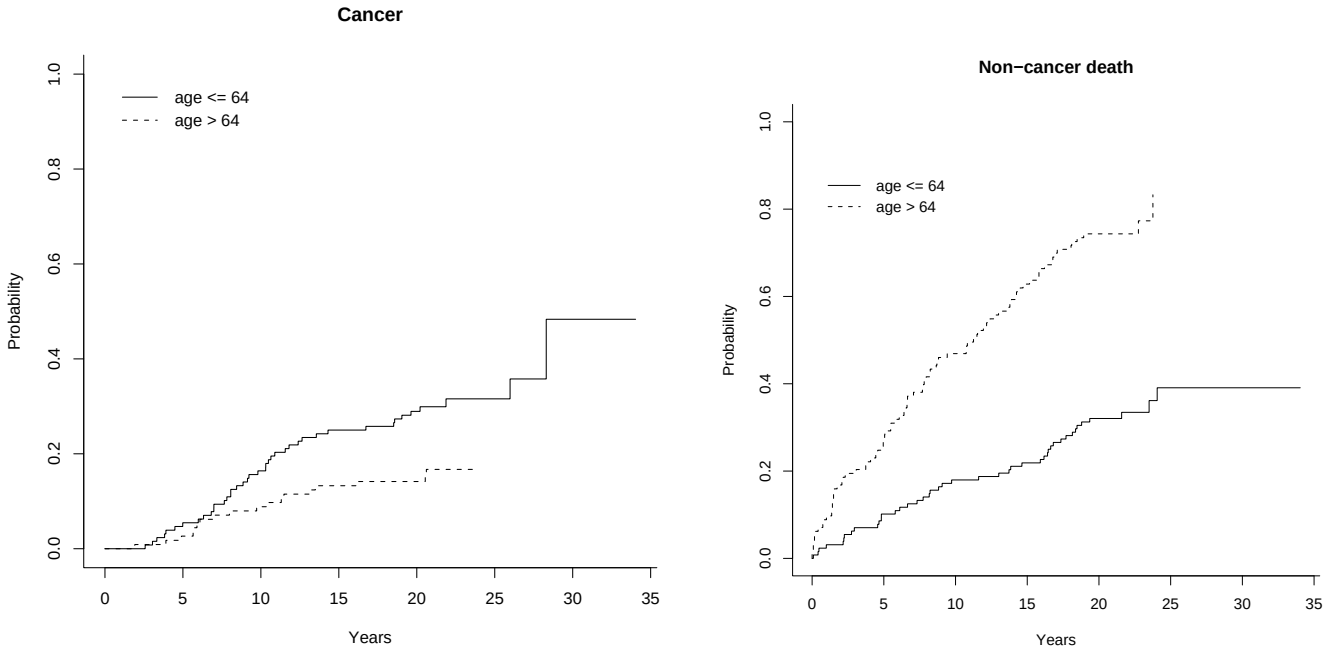
[a] Gray RJ. 1988, The Annals of Stat. 1988. 116: 1141–1154

Exemple : MGUS

- 241 patients with Monoclonal Gammopathy of Unknown Significance (MGUS)
- 20 to 35 years of follow-up
- Two competing events
  - Evolution towards a Cancer of the plasma cells ($\varepsilon = 1$) *Etc*
  - Death without transformation of the disease ($\varepsilon = 2$)
- 59 cancer of the plasma cells
- 130 death without transformation
- 52 censored patients

Kyle, Mayo Clinic Proceedings, 1993 and Grambsch & Therneau, 2000

Incidences cumulées : Age median 64 ans



Test de Gray : p=0.009 (EtC) , p=0.0001 (Décès)

Probabilité conditionnelle d'une evt concurrent

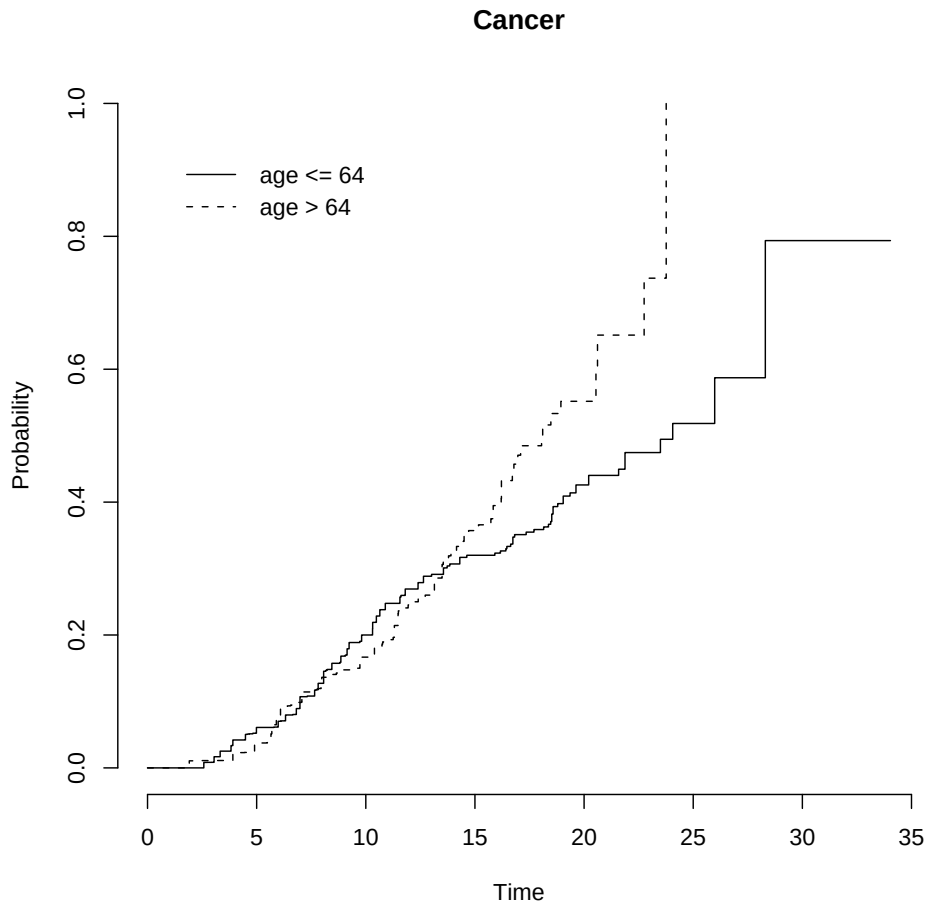
- Identifiable
- M.S. Pepe.Inference for Events With Dependent Risks in Multiple Endpoint Studies. 1991. *JASA*.1991
- M.S. Pepe and M. Mori. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? 1993. *Statistics in Medicine*, 12(8):737–751.
- A crude quantity that focuses on one risk type while **adjusting for other competing risks**
- In the first exposure, the CP was primarily a **descriptive device**

The CP of (Evol to) Cancer, CP_1 can be expressed as

$$\begin{aligned}\mathrm{CP}_1(t) &= P(\text{cancer by t}|\text{not dead from cancer by t}) \\ &= \frac{F_1(t)}{1 - F_2(t)}.\end{aligned}$$

Plugin estimator et test d'égalité de CP

Illustration



Pepe's test p=0.761

Modèle de regression proposé

The regression model for the conditional probabibility

$$\text{logit}(CP_1(t; \mathbf{Z}(t))) = \alpha(t) + \beta(t)'\mathbf{Z}(t) \qquad (1)$$

with $\alpha(t)$ is the (continuous) intercept, $\beta(t)$ is a vector of time–dependent regression coefficients and $Z(t)$ is a vector of time–dependent covariates.

Estimation : Temporal process regression

The temporal process regression[a] is a functional generalised linear model which specifies the **mean of a response** $Y(t)$ at time t conditionally on a vector of possibly time-dependent covariates $Z(t)$,

$$E(Y(t)|Z(t)) = g^{-1}(\beta(t)'Z(t))$$

Extension de Liang et Zeger (Longitudinal Data Analysis Using Generalized Linear Models, Biometrika.1986)

[a] Fine, Yan, Kosorok. Biometrika 2004

Estimation (1)

- Extend Marginal mean models, where all the covariate coefficients, are completely unspecified over time and therefore fully functional.
- At any fixed t, the TPR model is a generalized linear model.
- The time–varying coefficients at each time are estimated separately from estimating equations using the cross–sectional data at that time.

The cross–sectional data at each time are identified by some data–availability indicator $\delta(t)$

The information from the censoring time C is equivalent to a data availability indicator process constructed as $\delta(t) = I(t < C)$.

Suppose that, in a time window $[l, u]$, we observe n independent copies of $\{Y(t), Z, \delta(t)\}$, denoted $\{Y_i(t), Z_i, \delta_i(t)\}$

For estimating $\beta(t)$, we need

1. $E(Y(t)|Z) = E(Y(t)|Z, \delta(t) = 1)$ for each $t \in [l, u]$. $Y(t)$ is the same weather or not the data is observed $\leftrightarrow$ non-informative censoring in survival analysis.
2. The probability of observing complete data is positive $P(\delta(t) = 1|Z) > 0$ for each $t \in [l, u]$.
3. This assumption ensures at least some information for the cross-sectional data at each time point.

A 'working independence' estimator for $\beta(t)$, *i.e.* which does not take into account the temporal correlations, may be computed separately at each t.

Estimation (2)

Define $\hat{\beta}(t)$ as the root of

$$U(\beta(t), t) = \sum_{i=1}^{n} A_i(\beta(t), t),$$

where $A_i(\beta(t), t) = \delta_i(t) D_i'(\beta(t)) V_i(\beta(t), t)[Y_i(t) - g^{-1}(\beta(t) Z_i)]$, $D_i(\beta(t)) = \frac{d[g^{-1}(\beta(t) Z_i)]}{d\beta(t)}$ and $V_i(\beta(t), t)$ is a weight function

Let $l = \tau_0 < \tau_1 < \cdots < \tau_M = u$ be all the jump points observed in the data.

Then $\beta(t)$ only needs to be estimated at each of these jump points, and smoothing is not required.

Cas particulier : Odds proportionnel

Si on considère

- $g = logit$
- $Y(t) = I\{T \leq t, \varepsilon = 1 | T > t \cup \varepsilon = 1\}$

$$E(I\{T \leq t, \varepsilon = 1 | T > t \cup \varepsilon = 1\}) = CP_1(t)$$

$$\text{logit}(CP_1(t; \mathbf{Z}(t))) = \alpha(t) + \beta(t)'\mathbf{Z}(t)$$

$\rightarrow$ Sous modèle de TPR

Pour une covariable binaire on a

$$\frac{\text{CP}_1(t; Z)}{1 - \text{CP}_1(t; Z)} = e^{\beta(t)' Z} \frac{\text{CP}_{10}(t)}{1 - \text{CP}_{10}(t)},$$

où $\text{CP}_{10}(t) = \text{CP}_1(t; Z = 0)$ et $e^{\beta(t)}$ est l' OR.

Illustration : MGUS (2)

Comparaison avec les modèles classique (HR et SHR)

- Age binaire (age median 64)
- Age continu
- Taux de créatinine

Analyse univariés et β_j constants

EtC

Effect of age						
	HR	95% CI	SHR	95% CI	OR	95% CI
binary	0.85	[0.48 ; 1.49]	0.46⋆	[0.26 ; 0.80]	1.42	[0.69 ; 2.92]
continuous	1.00	[0.98 ; 1.03]	0.97	[0.96 ; 0.99]	1.03	[0.99 ; 1.06]
Effect of creatinine						
	0.86	[0.29 ; 2.60]	0.54	[0.28 ; 1.03]	1.04	[0.25 ; 4.26]

⋆

Death without prior evolution

Effect of age						
	HR	95% CI	SHR	95% CI	OR	95% CI
binary	3.82	[2.63 ; 5.55]	3.31	[2.44 ; 5.06]	6.41	[3.64 ; 11.30]
continuous	1.08	[1.06 ; 1.10]	1.08	[1.06 ; 1.10]	1.12	[1.09 ; 1.15]
Effect of creatinine						
	1.79	[1.38 ; 2.32]	1.80	[1.48 ; 2.19]	2.71	[1.15 ; 6.40]

Effets dépendant du temps

Pour une covariable binaire on a

$$\frac{\mathrm{CP}_1(t;Z)}{1-\mathrm{CP}_1(t;Z)} = e^{\beta(t)'Z}\frac{\mathrm{CP}_{10}(t)}{1-\mathrm{CP}_{10}(t)},$$

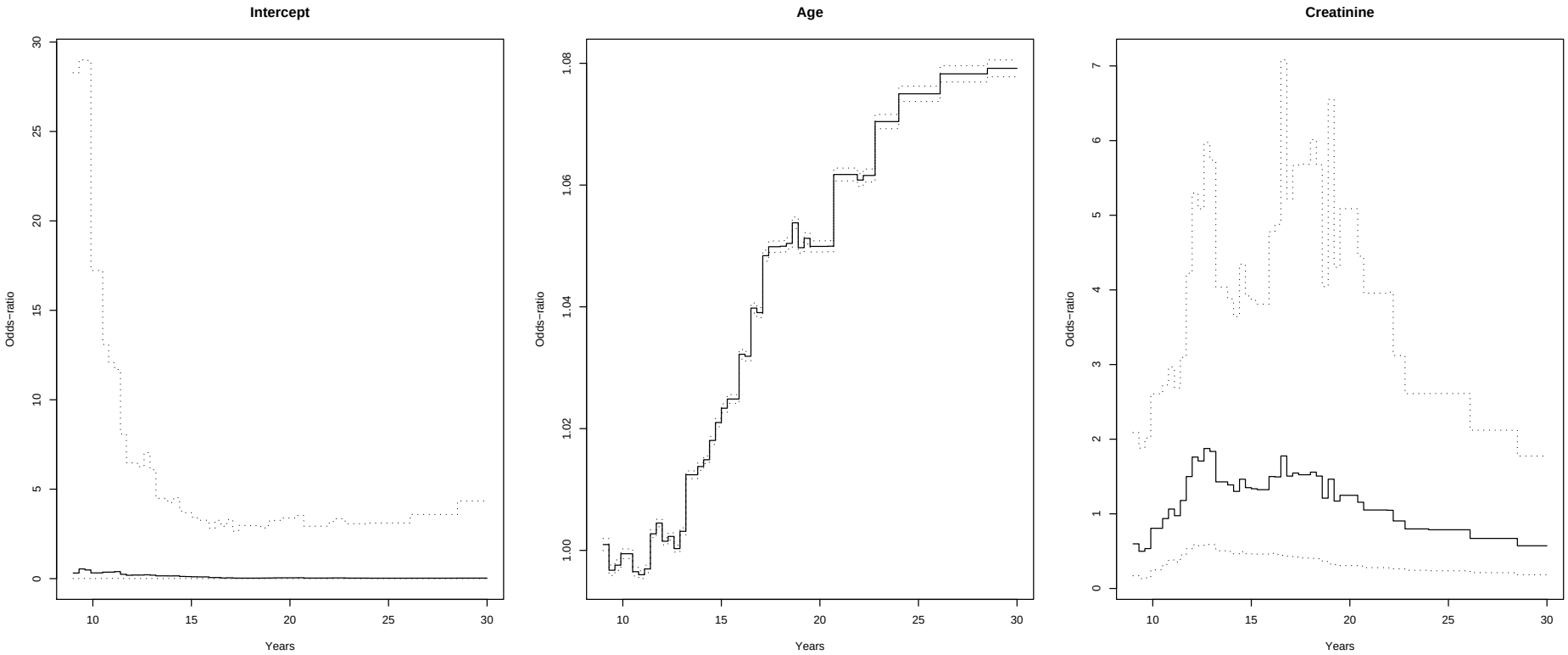
where $\mathrm{CP}_{10}(t) = \mathrm{CP}_1(t;Z=0)$ is the baseline CPF of EtC and $e^{\beta(t)}$ is a vector of odds-ratios.

Tests implémentés (Weighted integrated diff)

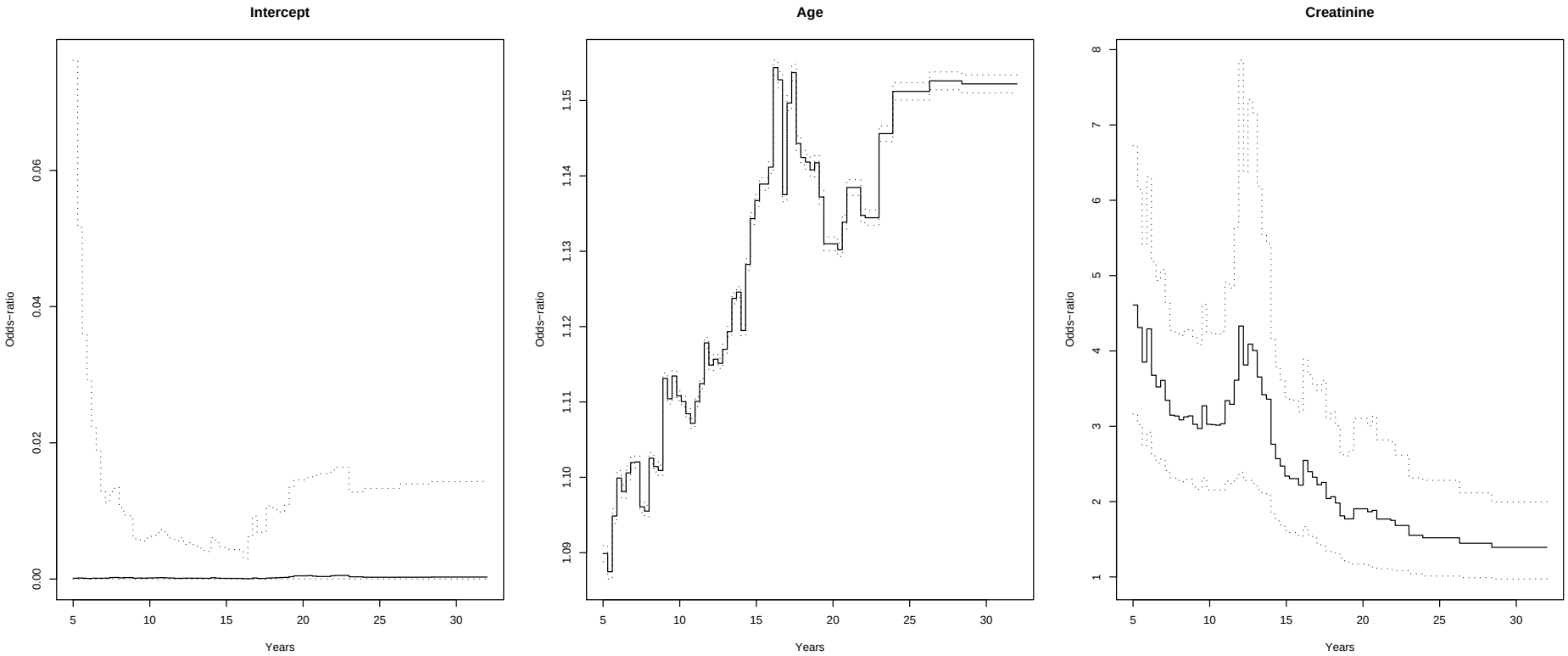
$H_0 : \beta_j(t) = 0$ ou $H_0 : \beta_j(t) = \beta_j$

Illustrations : Odd Ratio univariés

OR EtC


Intercept
Odds−ratio
0
5
10
15
20
25
30
10
15
20
25
30
Years
Age
Odds−ratio
1.00
1.02
1.04
1.06
1.08
10
15
20
25
30
Years
Creatinine
Odds−ratio
0
1
2
3
4
5
6
7
10
15
20
25
30
Years

OR Décès


Intercept
Odds−ratio
0.00
0.02
0.04
0.06
5
10
15
20
25
30
Years
Age
Odds−ratio
1.09
1.10
1.11
1.12
1.13
1.14
1.15
5
10
15
20
25
30
Years
Creatinine
Odds−ratio
1
2
3
4
5
6
7
8
5
10
15
20
25
30
Years

Effets dépendant du temps (2)

- Effet de l'age sur Death : dependant du temps ($p < 0,001$)
- Effet de la créatinine sur EtC est constant (p=0.94)
- Effet de la créatinine sur Death decroissant (p=0.32)

Conclusion

- Estimation dans un modèle à Odds proportionel avec effet dépendant du temps
- Covariables continues (mais nb limités)
- `Cprob` : Graphique, Test, Regression
- Interêt pour des covariables ayant des effets differents les CSH/CIF
- Rq: CPF saute quand un evt concurrent a lieu
- Covariables avec un effet important sur le Décès
- ToDo : GoF

Use of the **Cprob** package

The mgus data comes with the **Cprob** package, and can be loaded.

```
> data(mgus)
> head(mgus)
  id       time ev age creat
1  1 2.0807666  2  79   1.2
2  2 5.9137577  1  76   1.0
3  3 0.7583847  2  87   1.1
4  4 4.9691992  2  80   1.3
5  5 7.0828200  2  74   0.8
6  6 1.4976044  2  81   0.9
```

`id` stands for the patient identification number, `time` represents the time when an event occurs, and `ev` indicates the event type (0: censored observation, 1: Cancer, 2: Other). `age` and `creat` are the age and serum creatinine level at mgus diagnosis, respectively.

The `cpf` function estimates the CPF for the event of interest (specified through the `failcode` option). For example, this call provides the CPF of *EtC* for patients aged more or less than 64 years.

```
> mgus$A <- ifelse(mgus$age <= 64, 0, 1)
> cpf.dage.k <- cpf(mgus$time, mgus$ev, mgus$A, failcode=1)
```

The conditional probability curves can be obtained easily using

```
> plot(cpf.dage.k)
```

The `cpf.tpr` function is used to fit the proportional odds-model. This function relies on the **tpr** package. The following command fits the model for the CPF of *EtC* including age and creatinine.

```
> cov <- cbind(mgus$age, mgus$creat)
> tpr.mvc.k <- cpf.tpr(mgus$time, mgus$ev, cov, failcode=1,
+                      tis=seq(9,30,0.3), w=rep(1,91))
```

```
> tpr.mvc.k
Test for non-significant effect
                  stat     pvalue
(Intercept) -2.3434103 0.01910836
cov1         2.0703626 0.03841840
cov2         0.2158443 0.82910909

 Test for constant fit
                  coef   exp.coef    SE.coef       stat      pvalue
(Intercept) -3.08262971 0.04583856 1.31544599 -1.3897213 0.164613529
cov1         0.04050295 1.04133439 0.01956322  3.1235649 0.001786744
cov2         0.15896625 1.17229838 0.73648563 -0.1751134 0.860990557
```

The `print` method provides results from non-significant and constant fit tests.

Références

- Allignol A, Fine JP, Yan J, Latouche A. Regression Modelling of the Conditional Probability Function *submitted* 2008
- Survival Task Views : CRAN
- Beyersmann J, Latouche A, Buchholz A, Schumacher M. Simulating competing risks data in survival analysis. *Submitted* 2008
- Latouche A, Boisson V, Porcher R, Chevret S. Misspecified regression model for the subdistribution hazard of a competing risk. *Statistics in Medicine* 2007; **26**(5):965–974.

aurelien.latouche@uvsq.fr