Modelling of the effect of explanatory variables on health indicators: a pseudo-values approach

Camille SABATHÉ[1] - Pierre JOLY[1]

[1]Univ. Bordeaux, INSERM, Bordeaux Population Health, U1219, Bordeaux, France

GdR Statistiques et Santé 2017, Bordeaux - October 2017

Context

Models for lifetime risk of dementia accounting for interval censoring

Study of dementia:

- death is a competing risk
- interval censoring of the onset of the disease

Statistical methods need to account for this cohort data.

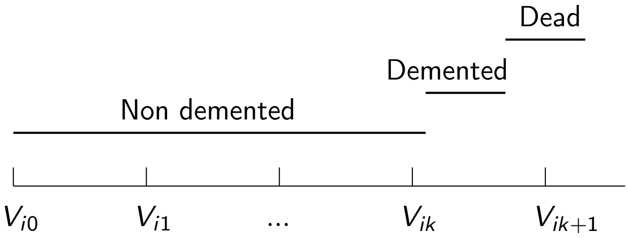
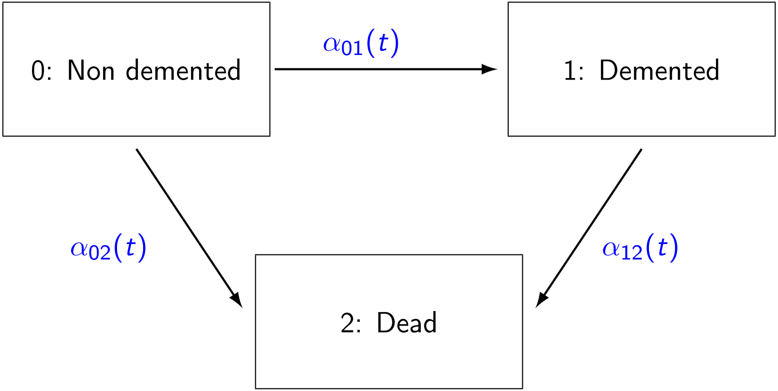


Figure 1: Example of follow-up of subject i

- diagnosis of dementia: onset of disease is between two date (V_{ik} and V_{ik+1})
- death without diagnosis: subject is demented ?

Illness-death model



with t the time and $\alpha_{01}(t)$, $\alpha_{02}(t)$ and $\alpha_{12}(t)$ the transition intensities

Health Indicators

$\alpha_{kl}(t)$: transition intensity from state k to state l
$A_{kl}(t) = \int_0^t \alpha_{kl}(u)\mathrm{d}u$

- Probability

$$P_{00}(t) = \exp\left[-A_{01}(t) - A_{02}(t)\right] \quad (1)$$

- Cumulative probabilities

$$F_{01}(t) = \int_0^t \exp\left[-A_{01}(u) - A_{02}(u)\right]\alpha_{01}(u)\mathrm{d}u \quad (2)$$

- Restricted mean survival time

$$RM(t) = \int_0^t \exp\left[-A_{01}(u) - A_{02}(u)\right]\mathrm{d}u \quad (3)$$

Inference for illness-death model

- proportional cause-specific hazards model
  - modelling of the transition intensities
  - $\alpha_{kl}(t \mid Z) = \alpha_{kl,0}(t) \exp(\beta_{kl}^T Z)$

$$F_{01}(t \mid Z) = \int_0^t \exp\left[-A_{01}(u)e^{(\beta_{01}^T Z)} - A_{02}(u)e^{(\beta_{02}^T Z)}\right] \alpha_{01}(u)e^{(\beta_{01}^T Z)} \mathrm{d}u$$

- Fine and Gray model
  - difficulty of interpretation
- Pseudo-values (Andersen et al., 2003)
  - developed with nonparametric estimator

Definition of pseudo-values

the pseudo-value $\hat{Y}_i$ for the subject i is defined by :

$$\hat{Y}_i(t) = n \times \hat{F}_{01}(t) - (n-1) \times \hat{F}_{01}^{-i}(t) \quad (4)$$

where $\hat{F}_{01}^{-i}$ is the leave-one-out estimator of $\hat{F}_{01}$ based on the sample of size n-1 without the ith subject and $\hat{F}_{01}$ the estimator of the cumulative incidence function

Pseudo-values are used for direct regression modelling of transition probabilities in a GLM with GEE.

- several link functions $g\,[.]$ (such as identity, log, cloglog) depending on the interpretation
- 1 time point / 5 to 10 time points (with or without interaction between time and covariates)

$$g\left[\mathbb{E}(\hat{Y}_i(t))\right] = \hat{\beta}_0 + \hat{\beta}_1 Z_i \tag{5}$$

$$g\left[\mathbb{E}(\hat{Y}_i(t))\right] = \hat{\beta}_0 + \hat{\beta}_1 Z_i + \hat{\beta}_2 t + \hat{\beta}_3 Z_i \times t \tag{6}$$

Pseudo-values with two types of estimators

$$\hat{Y}_i(t) = n \times \hat{F}_{01}(t) - (n-1) \times \hat{F}_{01}^{-i}(t)$$

Estimation of $\hat{F}_{01}(t)$ and $\hat{F}_{01}^{-i}(t)$ by:

- nonparametric approach:
  - n+1 Aalen-Johansen
  - R package `pseudo` (Klein et al.,2008)
- penalized likelihood with splines approach:
  - n+1 estimations of transition intensities models which are used to compute the n+1 estimator of $F_{01}(t)$
  - R package `SmoothHazard`

Design of simulations

n=500 with 500 replicates

Illness death model with the three transition intensities from Weibull distribution and $\alpha_{12}(t) < \alpha_{02}(t) < \alpha_{01}(t)$

1 covariate, $\mathbb{P}(Z = 1) = 0.4$ with a stronger effect of the covariate on death (regardless of disease status) than on disease

Pre-scheduled visit times were independently generated with consecutive visits spaced between 2 and 3 units of time.

True effect of Z in $F_{01}(t)$

$$\mathbb{E}\left[\hat{Y}_i(t)\right] = \hat{\beta}_0 + \hat{\beta}_1 \times Z_i$$

$$F_{01}(t = 18 \mid z = 0) = 0.334$$

$$F_{01}(t = 18 \mid z = 1) = 0.274$$

$$0.274 - 0.334 = -0.060$$

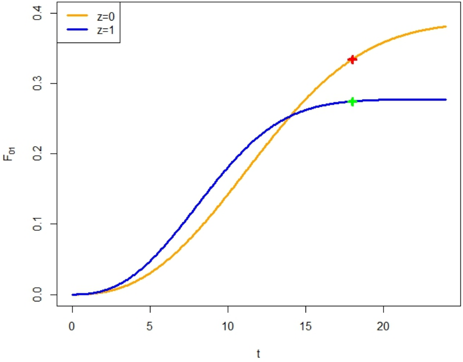


Table 1: Comparison of methods on the pseudo-values approach. Modelization of $\mathbb{E}\left[\hat{Y}_i(t)\right] = \hat{\beta}_0 + \hat{\beta}_1 Z_i$ with $t = 18$

β	meth.	$\overline{\hat{\beta}}$	Rel. Bias	RMSE $\times 1000$	Cov. rates
$\beta_0 = 0.334$	1	0.336	0.004	33	94.0
	2	0.332	-0.007	30	94.0
	3	0.273	-0.183	69	45.4
	4	0.337	0.007	38	91.6
$\beta_1 = -0.060$	1	-0.062	0.032	49	89.4
	2	-0.055	-0.071	47	94.2
	3	-0.058	-0.026	46	93.6
	4	-0.065	0.098	57	94.4

1 and 3: $\hat{F}_{01}(t)$ via Aalen-Johansen

2 and 4: $\hat{F}_{01}(t)$ via penalized likelihood and splines

results from interval censored data

Paquid Cohort

Study normal and pathological brain ageing, identify the prevalence, incidence, risk factors and preclinical manifestations of Alzheimer disease

Inclusion in 1988-1989 in Gironde and Dordogne
Participants had 65 years and over, not institutionalized
11 times of follow from 2 to 3 years over a 25-year period

n=3582
58% of women and 66% with medium or high educational level
age at baseline : 75.3 (sd=6.8)
MMS at baseline : 25.8 (sd=3.4)

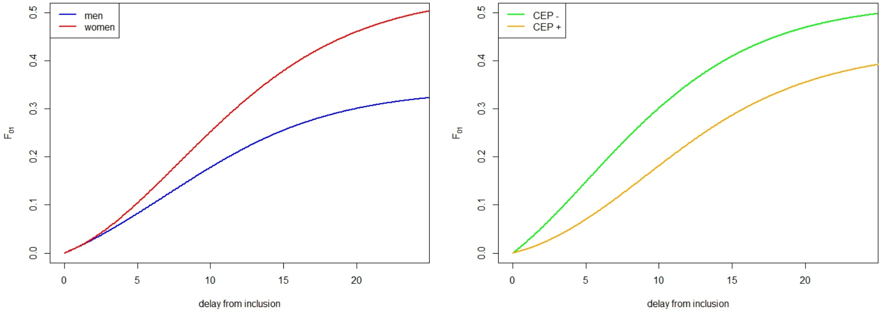


Figure 2: Cumulative probability of become demented from inclusion in Paquid Cohort by sex(left) and education level(right). n=3582

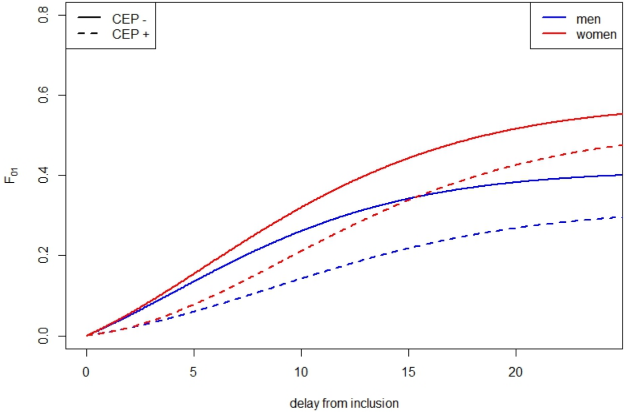


Figure 3: Cumulative probability of become demented from inclusion in Paquid Cohort by sex and education level. n=3582

Table 2: Results of pseudo-values approach on the probability of becoming demented until 10 years (ie. $\hat{F}_{01}(10)$) after inclusion for the effect of sex, education level, age at baseline and MMS

	$\hat{\beta}$	$CI_{95\%}$	p
man 65 yo. low educ. level and MMS=26	0.203	[0.159 ; 0.247]	<0.001
sex (women vs man)	0.041	[-0.002 ; 0.081]	0.038
education level (high vs low)	-0.016	[-0.064 ; 0.033]	0.531
age	0.010	[0.007 ; 0.013]	<0.001
MMS	-0.024	[-0.033 ;-0.015]	<0.001

Table 3: Results of pseudo-values approach on the probability of being alive and non-demented until 10 years (ie. $\hat{P}_{00}(10)$) after inclusion for the effect of sex, education level, age at baseline and MMS

	$\hat{\beta}$	$CI_{95\%}$	p
man 65 yo. low educ. level and MMS=26	0.397	[0.369 ; 0.425]	<0.001
sex (women vs man)	0.122	[0.096 ; 0.148]	<0.001
education level (high vs low)	0.036	[0.006 ; 0.065]	0.017
age	-0.030	[-0.032 ;-0.029]	<0.001
MMS	0.019	[0.015 ; 0.024]	<0.001

Table 4: Results of pseudo-values approach on expected time in state 0 until 10 years (ie. $\hat{RM}(10)$) after inclusion for the effect of sex, education level, age at baseline and MMS

	$\hat{\beta}$	$CI_{95\%}$	p
man 65 yo. low educ.level and MMS=26	7.074	[6.860 ; 7.287]	<0.001
sex (women vs man)	0.866	[0.676 ; 1.056]	<0.001
education level (high vs low)	-0.021	[-0.244 ; 0.203]	0.856
age	-0.177	[-0.192 ;-0.162]	<0.001
MMS	0.203	[0.166 ; 0.240]	<0.001

Discussion

Conclusion

Pseudo-values approach is a method to model directly the impact of explanatory variables on health indicators

Different possible interpretations according to the link functions and the fitted model

The method of estimation of the estimator affects computational time

Perspectives

Extension to left truncated data

Thank you for your attention

References :

Andersen, P. K., Klein J.P. et Rosthoj S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. Biometrika, 90(1), 15-27

Grand, M. K., et Putter, H. (2016). Regression models for expected length of stay. Stat Med, 35(7), 1178-92.

Joly, P., Commenges, D., Helmer, C., et Letenneur, L. (2002). A penalized likelihood approach for an illness-death model with interval-censored data: application to age-specific incidence of dementia. Biostat Oxf Engl, 3(3), 433-43

Klein, J. P., Gerster, M., Andersen, P. K., Tarima, S., et Perme, M. P. (2008). SAS and R functions to compute pseudo-values for censored data regression. Comput Methods and Programs in Biomed, 89(3), 289-300

Touraine, C., Gerds, T. A., et Joly, P. SmoothHazard: An R Package for Fitting Regression Models to Interval-Censored Observations of Illness-Death Models. J Stat Software. in press

	estimation	ic_inf	ic_sup	pval
(Intercept)	0.140	0.093	0.187	0.000
sex	0.042	0.003	0.081	0.035
education level	-0.017	-0.066	0.032	0.501
older vs younger	0.134	0.095	0.173	0.000
mms	-0.025	-0.034	-0.016	0.000

Table 5: F01 age binaire mms-26 lien id t=10

	estimation	ic_inf	ic_sup	pval
(Intercept)	0.565	0.533	0.598	0.000
sex	0.118	0.092	0.145	0.000
education level	0.039	0.009	0.069	0.012
age at baseline	-0.356	-0.383	-0.328	0.000
mms	0.025	0.021	0.029	0.000

Table 6: P00 age binaire mms-26 lien id t=10

	estimation	ic_inf	ic_sup	pval
(Intercept)	7.978	7.748	8.208	0.000
sex	0.842	0.648	1.037	0.000
education level	-0.002	-0.233	0.229	0.987
age at baseline	-1.909	-2.111	-1.708	0.000
mms	0.241	0.203	0.278	0.000

Table 7: RM age binaire mms-26 lien id t=10

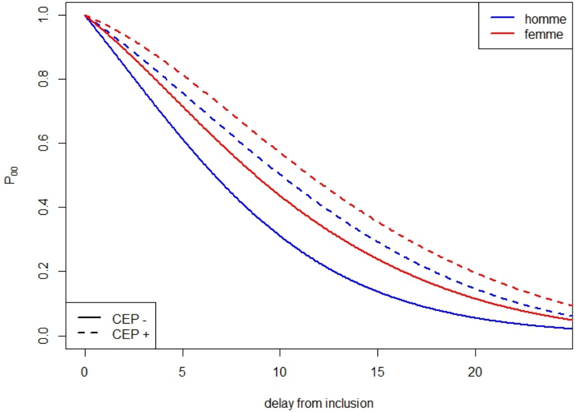


Figure 4: Probability of staying alive and non-demented from inclusion in Paquid Cohort by sex and education level. n=3582