Clustering of recurrent events data applied to the re-admission of elderly people at hospital

Vincent Vandewalle, Genia Babykina, Jean-Baptiste Beuscart, Jesus A. Carretero Bravo, Fabien Visade

Clustering of recurrent events data applied to the re-admission of elderly people at hospital

Vincent Vandewalle
Joint work with Génia Babykina, Jean-Baptiste Beuscart, Jesús Á. Carretero Bravo, Fabien Visade

European Conference on Data Analysis 2022
Thursday, September 15, 2022

Outline

Presentation of the cohort

DAMAGE cohort (Deschasse et al., 2021)

- Prospective multi-centric cohort of patients of 75 years old or more
- Hospitalised in France (Hauts de France and Normandy regions)
- 3081 patients of whom 1531 patients have had at least one re-hospitalisation
- Socio-demographic, health, geriatric variables

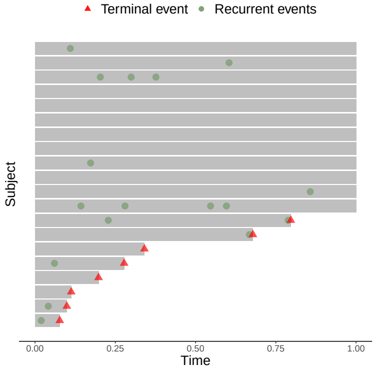
Recurrent event considered

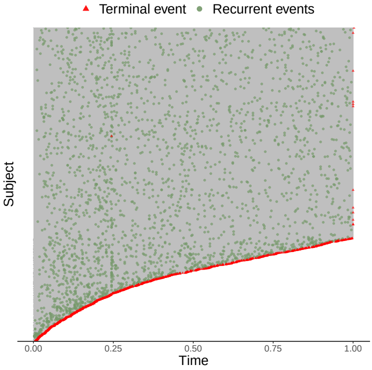
Successive readmission to the hospital

Medical questions

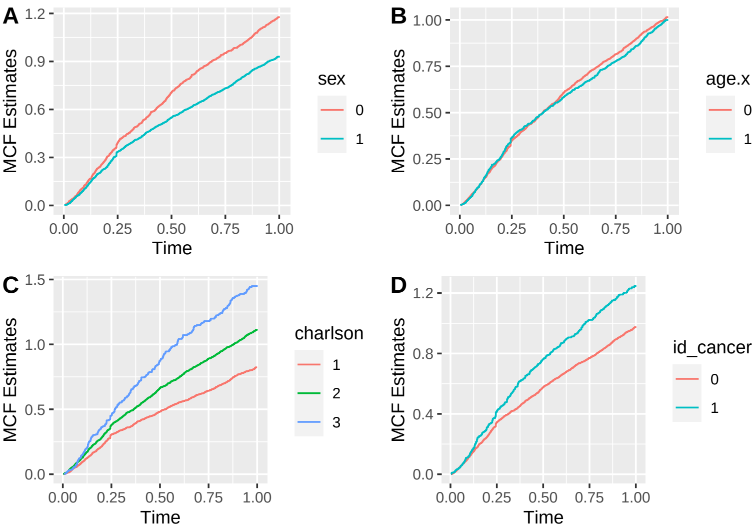
- Is the readmission process correlated with some clinical variables?
- Is the intensity of the readmission process correlated with survival?

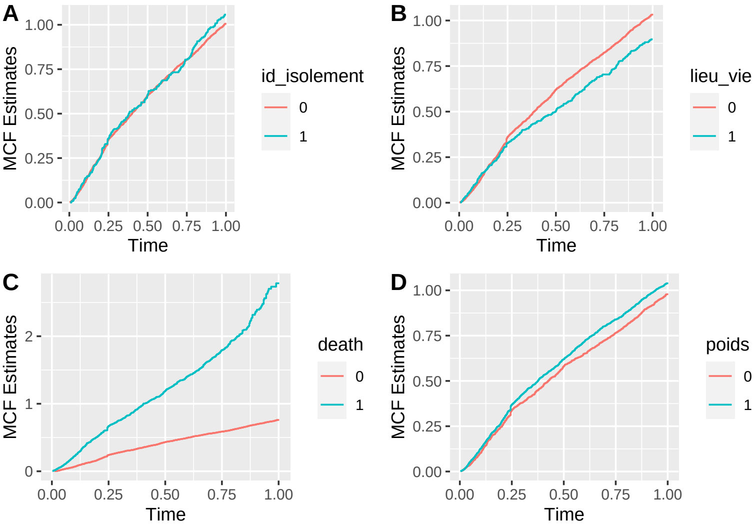
Visualization of recurrent events





Correlation of the event with other variables (1/2)




A
MCF Estimates
1.00
0.75
0.50
0.25
0.00
0.00 0.25 0.50 0.75 1.00
Time
id_isolement
0
1
B
MCF Estimates
Time
lieu_vie
0
1
C
MCF Estimates
2
1
0
Time
death
0
1
D
MCF Estimates
Time
poids
0
1

Goals of the proposed approach

Identify cluster of patients

- Adapt the treatment to each cluster
- Avoid therapeutic obstinacy

Assumptions

- Patients in a same cluster should follow the same recurrent events distribution
- The effect of covariates can be different in each cluster

Proposal

Assume that the data come from a mixture of counting processes with intensity depending on covariates

State of the art

Mixtures in survival analysis on time to first event: Cox and extensions

- Different censoring schemes (Peng, 2000)
- Censored data with high-dimensional covariates (Bussy et al., 2019)

Many applications in the analysis of recurrent events

Hospital readmissions, disease recurrences, repeated failures of industrial systems

Extensions of Cox model with recurrence: counting point process setting

- Andersen-Gill model (Andersen and Gill, 1982)
- Frailty models (Aalen, 1988; Oakes, 1992)

Mixtures of recurrent events

- Discrete-time duration model (McDonald and Rosina, 2001; Dean, Bauer, and Shanahan, 2014)
- Frailty-mixture models for recurrent event of *cure models* in health applications (Xu et al., 2012; Tawiah, McLachlan, and Ng, 2020).

Proposed model

Fully parametric mixture model based on the Andersen-Gill model with covariates

Notations

- $(T_{ij})_{j\geq 1}$: the times of successive events for i ($i = 1, \cdots, n$)
- $N_i(t)$ the number of events occurred to individual i
- $Z_i \in \{0,1\}^K$ the cluster of i in binary coding
- $\{N_i(t)\}_{t\geq 0}$ given $Z_{ik} = 1$ characterised by intensity $\lambda_{ik}(t)$

$$\lambda_{ik}(t)dt = \mathbb{E}\left[dN_i(t)|\mathcal{F}(t-), \boldsymbol{X}_i(t), Z_{ik} = 1\right],$$

$\boldsymbol{X}_i(t)$ the covariate process and $\mathcal{F}(t-)$ the filtration generated by the process

Intensity of the process in cluster k

$$\lambda_{ik}(t) = \lambda_{0k}(t)\exp(\boldsymbol{X}_i\boldsymbol{\beta}_k^{\intercal})$$

with baseline Weibull intensity $\lambda_{0k}(t) = \gamma_{1k}\gamma_{2k}t^{\gamma_{2k}-1}$

Parameters estimation

Likelihood of path i

$$p(\boldsymbol{t}_i|\boldsymbol{X}_i;\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \underbrace{\prod_{j=1}^{n_i} \left[\lambda_{ik}\left(t_{ij},\boldsymbol{\theta}_k\right)\right] \times \exp\left(-\int_0^{\tau_i} \lambda_{ik}(u,\boldsymbol{\theta}_k)du\right)}_{p(\boldsymbol{t}_i|\boldsymbol{X}_i,Z_{ik}=1;\boldsymbol{\theta})}.$$

EM algorithm

Starting from $\boldsymbol{\theta}^{(0)}$, alternate the following two steps until convergence

- **E-step** : update the class membership matrix at iteration h
  $\rho_{ik}^{(h)} = \frac{\pi_k^{(h-1)} p\left(\hat{\boldsymbol{\theta}}_k^{(h-1)}\right)}{\sum_{l=1}^{K} \pi_k^{(h-1)} p\left(\hat{\boldsymbol{\theta}}_k^{(h-1)}\right)}$,
- **M-step** : update model parameter by maximizing the expectation of the log-likelihood
  $\hat{\boldsymbol{\theta}}_k^{(h)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \sum_{i=1}^{n} \rho_{ik}^{(h)} \log p(\boldsymbol{t}_i|\boldsymbol{X}_i, Z_{ik} = 1;\boldsymbol{\theta})$

Selection of the number of clusters

BIC criterion

$$\mathsf{BIC}(K) = \ell(\hat{\boldsymbol{\theta}}; K) - \frac{\nu_K}{2} \log n$$

with $\nu_K = K(d+2)$, and n the number of paths

Remark

The choice of the covariates could also be performed using BIC

1. Re-admission of elderly people data

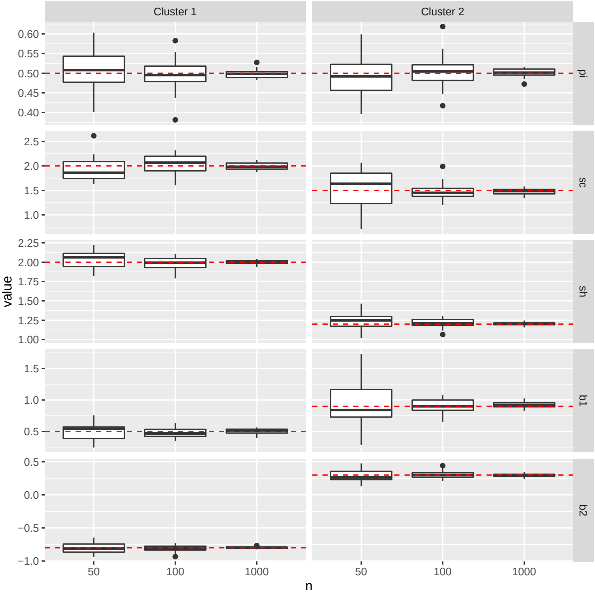
2. Model-based clustering of counting process

3. Results on simulated data

4. Results on re-admission data

5. Conclusion and perspectives

Consistency of the parameters estimation



- 20 replicates
- Bayes error: 14.5%
- Consistency of the parameters estimation

Output of the model

The variables sex and age are taken into account as covariates

BIC selects 2 clusters

Parameters of the clusters

	$\hat{\pi}_k$	$\hat{\gamma}_{1k}$	$\hat{\gamma}_{2k}$	$\hat{\beta}_{1k}$	$\hat{\beta}_{2k}$
Class 1	0.34	2.56	0.97	-0.16	-0.09
Class 2	0.66	0.58	0.85	-0.43	0.23

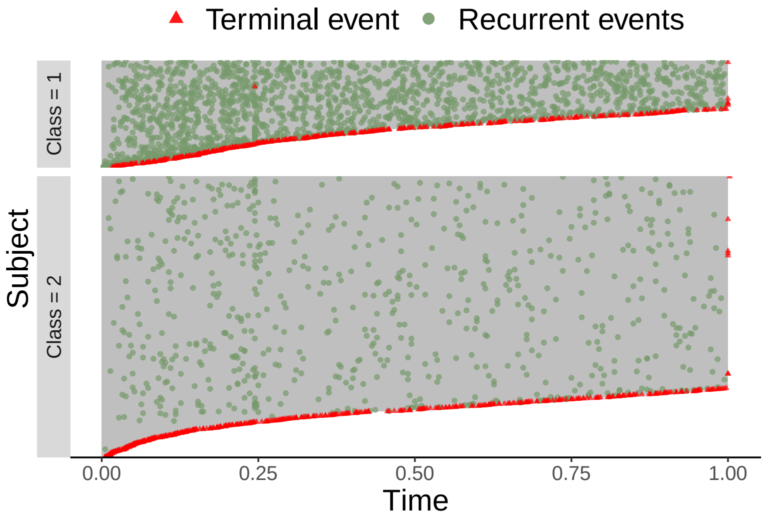
Distribution of the number of events given to the cluster (%)

	1	2	3	4	5	6	7	8	9
Class 1	0	29	44	16	7.3	1.9	1.2	0.6	0.2
Class 2	70	30	0.1	0	0	0	0	0	0

Interpretation with respect to other covariates

	age.x	sex	death
Class 1	30%	63%	54%
Class 2	31%	67%	25%

Visualisation of the clusters



Conclusion and perspectives

Conclusion

- Able to detect different recurrence types in the data
- Not uniquely based on the number of event but also on the variation of the intensity
- Additional interpretation of the re-admission process

Perspectives

- Make depend the prior probability on baseline covariates
- Perform prognosis of the patient cluster
- Study how to integrate the information on the patient cluster at baseline to medical practices

References I

- Aalen, Odd O (1988). "Heterogeneity in survival analysis". In: *Statistics in medicine* 7.11, pp. 1121–1137.
- Andersen, Per Kragh and Richard D Gill (1982). "Cox's regression model for counting processes: a large sample study". In: *The annals of statistics*, pp. 1100–1120.
- Bussy, Simon et al. (2019). "C-mix: A high-dimensional mixture model for censored durations, with applications to genetic data". In: *Statistical methods in medical research* 28.5, pp. 1523–1539.
- Dean, Danielle O, Daniel J Bauer, and Michael J Shanahan (2014). "A discrete-time Multiple Event Process Survival Mixture (MEPSUM) model.". In: *Psychological methods* 19.2, p. 251.
- Deschasse, G. et al. (2021). "Development of a predictive score for mortality at 3- and 12-month after discharge from an acute geriatric unit as a trigger for advanced care planning". In.
- McDonald, John W and Alessandro Rosina (2001). "Mixture modelling of recurrent event times with long-term survivors: Analysis of Hutterite birth intervals". In: *Statistical Methods and Applications* 10.1, pp. 257–272.

References II

Oakes, David (1992). "Frailty models for multiple event times". In: *Survival analysis: State of the art*. Springer, pp. 371–379.

Peng, Fengchun (2000). "A mixture-model approach to the analysis of survival data". In: *Generalized Linear Models: A Bayesian Perspective*. CRC Press. Chap. 15.

Tawiah, Richard, Geoffrey J McLachlan, and Shu Kay Ng (2020). "Mixture cure models with time-varying and multilevel frailties for recurrent event data". In: *Statistical methods in medical research* 29.5, pp. 1368–1385.

Xu, Ying et al. (2012). "Estimation of summary protective efficacy using a frailty mixture model for recurrent event time data". In: *Statistics in medicine* 31.29, pp. 4023–4039.