MOX-Report No. 25/2017

Joint modelling of recurrent events and survival: a Bayesian nonparametric approach

Paulon, G.; De Iorio, M.; Guglielmi, A.; Ieva, F.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it http://mox.polimi.it

Joint modelling of recurrent events and survival: a Bayesian nonparametric approach

GIORGIO PAULON*,
University of Texas at Austin (USA)
giorgio.paulon@utexas.edu

MARIA DE IORIO,
Department of Statistical Science, University College London, UK
ALESSANDRA GUGLIELMI, FRANCESCA IEVA
Dipartimento di Matematica, Politecnico di Milano, Italy

May 6, 2017

Abstract

Heart failure (HF) is one of the main causes of morbidity, hospitalization and death in the western world and the economic burden associated with HF management is relevant and expected to increase in the future. We consider hospitalization data for heart failure in the most populated Italian Region, Lombardia. Data were extracted from the administrative data warehouse of the regional healthcare system. The main clinical outcome of interest is time to death and research focus is on investigating how recurrent hospitalizations affect the time to event. The main contribution of the paper is to develop a joint model for gap times between two consecutive hospitalizations and survival time. The probability models for the gap times and for the survival outcome share a common patient specific frailty term. Using a Bayesian nonparametric prior as the random effects distribution accounts for patient heterogeneity in recurrent event trajectories. Moreover, the joint model allows for dependent censoring of gap times by death or administrative reasons and for the correlations between different gap times for the same individual. It is straightforward to include covariates in the survival and/or recurrence process through the specification of

appropriate regression terms. Posterior inference is performed through Markov chain Monte Carlo methods.

AFT model, Dirichlet process mixtures, frailty, survival analysis, waiting times.

1 Introduction

Recurrent event data arise in different fields of application and typical examples include engineering product testing and reliability analysis of repairable systems in technology, recurrent infections and hospitalizations in medicine. In this article we focus on medical applications. In this context it is typical to have observations on a large number of individuals, each of them presenting a small number of occurrences of the clinical event of interest. In many applications, a terminating event such as death can occur during the follow-up period precluding further occurrence of the recurrent events. When recurrent event processes are terminated by another absorbing event, data are usually referred to as recurrent events data with termination and relevant information may include patients' survival times. The termination (often death) time may be dependent on the recurrent event history and it is essential to account for dependence between the recurrent and terminal event processes; for example, Schmoor et al. (2013) and Conlon et al. (2014) use cancer relapses to predict the risk of death.

The main contribution of this work is to develop a joint model for waiting times between recurrent events and survival outcome within a Bayesian nonparametric framework. Particular attention is devoted to the clustering of subjects according to their history of recurrent events and termination and the ability of assessing the relationship between event occurrence, survival and potential explanatory factors. We treat the time-to-event as the main clinical outcome of interest and we model the relationship between survival times and recurrence of events. This is an important feature of our approach as several recurrences (in our case re-hospitalizations) are often related to risk of death and are likely to affect it. The strength of the association between recurrences and terminal event may then be interpreted in terms of patients' risk profiling, and a better understanding of how recurrences affect survival may lead to a more effective planning of healthcare resources. In this sense, the terminal event can be considered as informative censoring.

To better accommodate for subject-specific variability in the recurrent event trajectories, we specify a random effect distribution based on the Dirichlet Process Mixture

prior. This choice allows for extra flexibility, over-dispersion and clustering of the observations and overcomes the often too restrictive assumptions underlying a parametric distribution. Similarly to Huang and Liu (2007), we model the time dependency between recurrent events assuming that, conditional to subject-specific random effects parameters, the gap (or waiting) times between such events are independent. We then assume that the conditional distribution of the survival time for each individual depends on the same random effect parameters. In other words both conditional distributions, i.e., the one of the j-th gap times and the one of the survival time, share a common subject-specific frailty, that is a subject-specific random effect on the log scale. The joint model takes into account the dependent "censoring" of gap times by death, and the dependency between different gap times of the same patient. However, unlike Huang and Liu (2007), we model each event time distribution as a regression model, and our approach is Bayesian. The shared frailty parameters are given a Bayesian nonparametric prior, in particular they are a sample from a Dirichlet process (DP) (Ferguson, 1973). It is well known that the DP is almost surely discrete, and that if G is a DP(M, G_0) with total mass parameter M and baseline distribution G_0, then G can be represented as (Sethuraman, 1994)

$$G(\cdot) = \sum_{h \geq 1} w_h \delta_{\theta_h}(\cdot) \tag{1}$$

where δ_θ is a point-mass at θ, the weights follow a stick-breaking process, $w_h = V_h \prod_{j<h}(1 - V_j)$, with $V_h \overset{\text{iid}}{\sim} \text{Beta}(1, M)$, and the atoms $\{\theta_h\}_{h \geq 1}$ are such that $\theta_h \overset{\text{iid}}{\sim} G_0$. Due to the discreteness of the DP, the prior induces clustering of the subjects in the sample based on the unique values of the random effects parameters, where the number K of clusters is unknown and learnt from the data. Brown and Ibrahim (2003) use a similar strategy for specifying a joint model for survival and longitudinal outcome to allow for extra flexibility and robustness in the model. Ouyang et al. (2013) develop statistical methods for joint modelling of recurrent event counts and survival time for heart transplantation patients within a Bayesian framework, where the emphasis is on modelling the risk of death and the risks of rejections. See Sinha et al. (2008) for a detailed review of such approaches. In a frequentist framework, Yu and Liu (2011) model nonparametric covariate functions in the presence of recurrent events and dependent termination.

In summary, our strategy consists in specifying a survival regression model for the time-to-event response and a distribution for the gap times between recurrent

events. An alternative approach consists in modelling the intensity functions of the event counts of the recurrent process, and the survival hazard rate. For a discussion about the relative merits of the two approaches see Cook and Lawless (2007).

Our work is motivated by a real data application involving Congestive Heart Hailure (hereafter HF for sake of simplicity) patients. HF is a chronic disease caused by many conditions that damage the heart muscle, including coronary artery disease, heart attack, cardiomyopathy and conditions that overwork the heart (high blood pressure, valve disease, thyroid disease, kidney disease, diabetes or heart defects present at birth). In addition, HF can occur due to a combination of these diseases. This morbid illness associated to a very poor prognosis, often leading to death or repeated hospitalizations, which are both largely burdensome to the patient and the healthcare system. For instance, it has been estimated that the average cost of a HF-related event in Lombardia is around 6,000 euros. Despite the efforts to improve the efficiency and the efficacy of treatments and management, re-hospitalization rates remain persistently high. Moreover, the ageing of the population and improved survival of cardiac patients due to modern therapeutic innovations have led to an increasing impact of HF on healthcare systems all over the western countries. As for many other chronic diseases, clinical interest lies in both the final outcome (death or survival time) and the dynamics of the process itself, since it determines the subsequent quality of patients' life. From an economic and healthcare planning perspective, there is great interest in strategies to reduce re-hospitalization for HF. In fact, a better understanding of both death and non-fatal clinical events would potentially lead to improved prognosis and a better assessment of the impact and costs of the disease by healthcare providers. It is therefore paramount to develop a comprehensive model for disease management, mortality and associated clinical event histories, which is also able to account for the significant inter-individual variability in disease course which is typical of chronic diseases, as well as of biological events. In our analysis we use episodes of hospitalization for disease related events (recurrent events), obtained from administrative healthcare data of Lombardia (one of the administrative divisions of Italy), and we include patients characteristics to predict risk of death.

In Section 2 we introduce the model, while in Section 3 we describe in details the application. In Section 4 posterior inference results are presented, then in Section 5 we assess model performance and goodness of fit and in Section 6 we discuss the proposed approach in terms of out-of-sample predictive ability. Eventually, in 7 we compare our

approach to other competing methods. We conclude the paper in Section 8.

2 A joint model for gap times of recurrent events with termination

We consider data on N individuals. We assume that $0 := T_{i0}$ corresponds to the start of the event process for individual i and that subject i is observed over the time interval $[0, \zeta_i]$. If n_i events are observed at times $0 < T_{i1} < \cdots < T_{in_i} < \zeta_i$, let $W_{ij} = T_{ij} - T_{ij-1}$ for $j = 1, \ldots, n_i$ denote the waiting times (gap times) between events of subject i. Let S_i denote the survival time of patient i since the start of the corresponding event process: either the time S_i or the censoring time ζ_i is observed. If S_i is observed, then $\zeta_i = S_i$ and $T_{in_i} < S_i$, otherwise $T_{in_i} < \zeta_i$. In what follows we set $T_{in_i+1} > \zeta_i$ with the last gap-time $T_{in_i+1} - T_{in_i}$ always censored. Let J be the maximum number of observed repeated events, i.e., $J := \max_{i=1,\ldots,N} n_i$.

As mentioned in the Introduction, our goal is to jointly model the gap times and survival time of each subject in the sample. Two approaches are possible. One consists of introducing an explicit dependence of the survival time on an underlying process, as for instance in Brown and Ibrahim (2003), who specify a Bayesian nonparametric model for a longitudinal process whose outcome (i.e., the trajectory) is used as predictor in the hazard function of the time-to-event. The second strategy consists of assuming that, conditionally on all the parameters, gap times are independent of each other and are also independent of the survival time. In this latter case, the shared parameters take into account the dependent "censoring" of gap times by termination and the correlation between different gap times for the same subject. Huang and Liu (2007) and Ouyang et al. (2013) use a common frailty parameter in order to link the two hazard functions of the waiting times and of the time-to-event. In this paper, we opt for the second strategy and develop a Bayesian semiparametric approach.

Since the terminal event censors event recurrence, but not vice versa, we need to assume a semi-competing risks model, i.e., a model taking into account that, when subjects are at risk of another recurrent event, they are also at risk of the terminal event; see, for instance Cook and Lawless (2007, Sect. 6.6). More specifically, let $N_i(t)$ and $\Delta N_i(t)$ be the number of recurrent events on the interval $[0, t]$ and $[t, t + \Delta t)$, respectively. We assume $N_i(0) = 0$ and define $D_i(t) := \mathbf{1}(t \leq S_i)$. Let $H_i(t) := \{(N_i(s), D_i(s)) : 0 \leq s < t\}$ be the process history of subject i up to time t. The

intensity function of the recurrent process is given by

$$\lambda_i(t|H_i(t)) = \lim_{\Delta t \downarrow 0} \frac{P(\Delta N_i(t) = 1|H_i(t))}{\Delta t}$$

while the intensity function of the terminal event is

$$\gamma_i(t|H_i(t)) = \lim_{\Delta t \downarrow 0} \frac{P(S_i < t + \Delta t|H_i(t), D_i(t) = 1)}{\Delta t} \ .$$

We assume that

$$\begin{aligned} \lambda_i(t|H_i(t)) &= \lambda_i(t|H_i(t), x_i(t)) = h_i(t - T_{N_i(t^-)}|x_i(t)) \\ \gamma_i(t|H_i(t)) &= \gamma_i(t|H_i(t), x_i(0)) \end{aligned} \tag{2}$$

where $t - T_{N_i(t^-)}$ is the time since the most recent event before t for subject i, $x_i(t)$ is a vector of covariates at time t and $h_i(\cdot)$ is the hazard function of gap times W_{ij} (see below for more modelling assumptions). We introduce dependency among gap times and between gap times and the terminal event time of the same patient through random effects. However, in contrast to Huang and Liu (2007), instead of explicitly modeling the intensities or the hazard of the recurrent event process, we assume an accelerated failure time model with random effects linking gap and terminal event times.

Specifically, we specify the following hierarchical structure for the log-transformation of waiting times and survival times, i.e., $Y_{ij} = \log(W_{ij})$, $j = 1, \ldots, n_i+1$, $U_i := \log(S_i)$, $i = 1, \ldots, N$:

$$Y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\beta}_j^*, \alpha_i, \sigma_i^2 \overset{\text{ind}}{\sim} \mathcal{N}(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_j^* + \alpha_i, \sigma_i^2) \quad j = 1, \ldots, n_i + 1, \tag{3}$$

$$U_i|\boldsymbol{z}_i, \boldsymbol{\gamma}, \alpha_i, \psi, \eta_i^2 \sim \mathcal{N}(\boldsymbol{z}_i^T\boldsymbol{\gamma} + \psi\alpha_i, \eta_i^2). \tag{4}$$

for $i = 1, \ldots, N$. Therefore, $(Y_{i1}, \ldots, Y_{in_i+1})$ and U_i are conditionally independent for each i, given the hyper-parameters, as well as the trajectories for different patients. Here $\boldsymbol{\beta}_j^* := (\boldsymbol{\beta}_0, \boldsymbol{\beta}_j)^T = (\beta_{01}, \ldots, \beta_{0p}, \beta_{j1}, \ldots, \beta_{jq})^T$ is the vector of regression coefficients, $\boldsymbol{x}_{ij}$ is a set of p fixed and q time-varying covariates influencing the gap times, while $\boldsymbol{\gamma} := (\gamma_1, \ldots, \gamma_r)$ and $\boldsymbol{z}_i$ denote the vector of regression coefficients and fixed covariates, respectively, which are potential predictors of the time-to-event. Note that treatment effects on disease recurrence and survival are not necessarily the same, since, in general, some therapies may delay disease recurrence but not prolong survival. For this reason, the covariates $\boldsymbol{z}_i$ and the components of $\boldsymbol{x}_{ij}$ may be distinct. Observe that $h_i(\cdot) = h_i(\cdot|\boldsymbol{x}_{ij}, \boldsymbol{\beta}_j^*, \alpha_i, \sigma_i^2)$ in (2) is defined in our model as the hazard of W_{ij} from the log-normal distribution in (3), while, similarly, $\gamma_i(\cdot) = \gamma_i(\cdot|\boldsymbol{z}_i, \boldsymbol{\gamma}, \alpha_i, \psi, \eta_i^2)$ is the

hazard derived from the distribution of S_i in (4). The likelihood for subject i, under independent censoring, is then given by:

$$\left(\prod_{j=1}^{n_i} f_Y(y_{ij}|\boldsymbol{x}_{ij},\boldsymbol{\beta}_0,\boldsymbol{\beta}_j,\alpha_i,\sigma_i^2)\right) S_Y(\log(\tau_i-(\mathrm{e}^{y_{i1}}+\ldots+\mathrm{e}^{y_{in_i}})))$$
$$\times f_U^{1-\nu_i}(\log\tau_i|\boldsymbol{z}_i,\boldsymbol{\gamma},\alpha_{ij},\eta^2) S_U^{\nu_i}(\log\tau_i|\boldsymbol{z}_i,\boldsymbol{\gamma},\alpha_{ij},\eta_i^2),$$

where f_Y, f_U are the densities of the gap and survival times (both Gaussian), respectively, S_Y, S_U denote the corresponding survival functions, $\tau_i = \min(S_i, \zeta_i)$ and ν_i $(= D_i(\zeta_i))$ is the censoring indicator, which is equal to 1 if the survival time is censored and 0 otherwise. Note that the factor $S_Y(\log(\tau_i-(\mathrm{e}^{y_{i1}}+\ldots+\mathrm{e}^{y_{in_i}})))$ is the contribution of the n_i+1-th gap time being censored, i.e., the conditional probability that W_{in_i+1} is larger than $\tau_i-(W_{i1}+\ldots+W_{in_i})$, given parameters and past history. As stated earlier, given the parameters and the covariates, the individual recurrent processes are assumed conditionally independent. We are making the implicit assumption (which is common in these type problems) that a patient cannot experience a recurrent event and a terminal event at the time. However, in our framework, this assumption is easily relaxed.

We assume a priori independence among parameters $\boldsymbol{\beta}_0$, $(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_J)$, $\boldsymbol{\gamma}$, ψ, and $\{(\alpha_i,\sigma_i^2,\eta_i^2),\ i=1\ldots,N\}$. As random effect distribution we specify a nonparametric prior distribution for $(\alpha_i,\sigma_i^2,\eta_i^2)$:

$$\begin{aligned}&(\alpha_i,\sigma_i^2,\eta_i^2)|G \overset{\text{iid}}{\sim} G \quad i=1,\ldots,N\\ &G\sim \mathrm{DP}(M,G_0),\end{aligned} \tag{5}$$

i.e., the random effects distribution is a Dirichlet Process.

In summary, our modelling assumptions imply that (i) waiting times are independent of each other, conditionally to the other parameters; (ii) the subject-specific random effect α_i links the distribution of the waiting times and survival times, as it determines the mean of the distribution of Y_{ij} and it is used as predictor in the survival regression component of the likelihood (see (4)); (iii) the shared parameter α_i allows the clustering to depend on both gap times trajectories and survival outcome. We complete the model by setting the following prior distribution on the remaining

parameters:

$$
\begin{aligned}
&\boldsymbol{\beta}_0 \sim \mathcal{N}_p(\mathbf{0}, \beta_0^2 I_p) \\
&\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J \mid \boldsymbol{\mu} := (\mu_1 \ldots, \mu_q)^T, \Sigma := \text{diag}(\tau_1^2, \ldots, \tau_q^2) \overset{\text{iid}}{\sim} \mathcal{N}_q(\boldsymbol{\mu}, \Sigma) \\
&\mu_1, \ldots, \mu_q \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\mu^2) \\
&\tau_1^2, \ldots, \tau_q^2 \overset{\text{iid}}{\sim} \text{Inv-Gamma}(a_\tau, b_\tau) \\
&\boldsymbol{\gamma} \sim \mathcal{N}_r(\mathbf{0}, \gamma_0^2 I_r) \\
&\psi \sim \mathcal{N}(0, \psi_0^2) \\
&G_0 = \mathcal{N}(0, \alpha_0^2) \times \text{inv-Gamma}(a_\sigma, b_\sigma) \times \text{inv-Gamma}(a_\eta, b_\eta) \\
&M \sim \mathcal{U}(a_M, b_M).
\end{aligned}
\tag{6}
$$

Note the use of a further level of hierarchy in the prior marginal distribution of the time-varying regression coefficients to ensure exchangeability. This allows the coefficients to exchange information over time and leads to better estimates, in particular for the last gap times as often fewer observations are available.

The model can be generalized by making different distributional assumptions for either/both gap and survival times (e.g. using a Weibull distribution), or by allowing the variance of the gap times to depend on the time index. As an alternative, we could assume $\eta_i = \eta$ for $i = 1, \ldots, N$, and a parametric marginal prior for it, i.e., $\eta^2 \sim$ inv-Gamma(a_η, b_η), while maintaining the nonparametric prior specification (α_i, σ_i^2):

$$
\begin{aligned}
&(\alpha_i, \sigma_i^2) | G \overset{\text{iid}}{\sim} G \quad i = 1, \ldots, N \\
&G \sim \text{DP}(M, G_0) \\
&G_0 = \mathcal{N}(0, \alpha_0^2) \times \text{inv-Gamma}(a_\sigma, b_\sigma).
\end{aligned}
\tag{7}
$$

Finally, it is easy to perform variable selection in this context, by assuming, for example, a spike and slab prior on the regression coefficients or performing Stochastic Search Variable Selection. See Rockova et al. (2012) for a review of Bayesian variable selection strategies.

3 Congestive heart failure dataset

We apply the model described in Section 2 to a real dataset extracted from the health-care data warehouse of Regione Lombardia (a region in Northern Italy), which contains information on patient healthcare usage and the relative economic impact on the national health system (e.g. hospitalizations, drugs, visits); see Mazzali et al. (2015) and

Mazzali et al. (2016) for details. We consider data on a sample of $n = 1000$ patients coming from the dataset described in Mazzali et al. (2016). The subsample is representative of the entire population in terms of age, gender, comorbidity burden, number of procedures and groups. We focus on hospitalizations due to Congestive Heart Failure (HF) in the time window January 1st, 2006 - December 31th, 2012. Therefore, for administrative reasons, the censoring time for all the patients in the sample is December 31th, 2012. In the analysis the gap times refer to times between successive hospitalizations. The first recorded hospitalization for each patient represents here the origin of the recurrent process ($T_{i0} := 0$ for all i); consequently, n_i represents the number of completely observed gap times between subsequent hospitalizations, given the initial one.

Since the number of recurrences for patients differs widely, we only consider patients with at least two recurrences (including the first one), i.e. at least one observed gap time but no more than 10. The resulting dataset for the analysis consists of $N = 810$ patients for a total of 2920 gap times (this subset covers 74.64% of all the events). Table 1 reports the distribution of the number patients N_j for which j , $j = 1, \ldots, J = 10$ waiting times are observed, where $\sum_j N_j = N$.

j	1	2	3	4	5	6	7	8	9	10	TOT
N_j	169	153	142	100	71	66	50	24	24	11	810

Table 1: Number of patients N_j that experience exactly j recurrent events, $j = 1, \ldots, J$.

Figure 1 displays the histogram of the observed gap times in log-scale: 356 out of 810 patients are right-censored (in terms of event time), which corresponds to a high censoring rate, approximately 44%. In Figure 2 we show the empirical distribution of event times for censored (blue) and non-censored (red) observations in log scale. This implies that for each patient the likelihood contribution from the waiting time process includes always a last censored waiting time, independently of censoring by death or administrative reasons.

Information has also been collected on several covariates, fixed or time-varying. We report below a list of the covariates included in the model:

- *gender* of the patient. In Table 2, we report the percentage p_j of male patients among the observations available at each gap time j. This proportion is roughly close to 50% except for the last one.

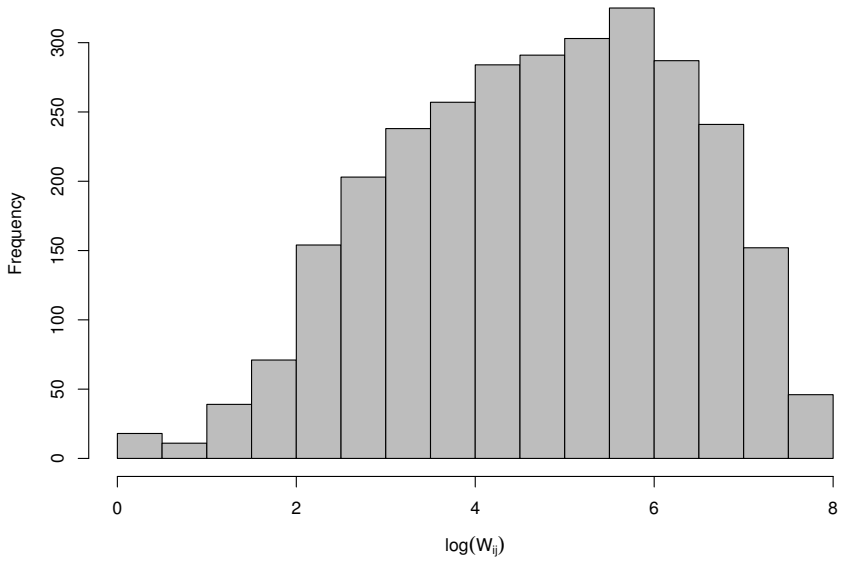


Figure 1: Histogram of the log-transformed gap times.

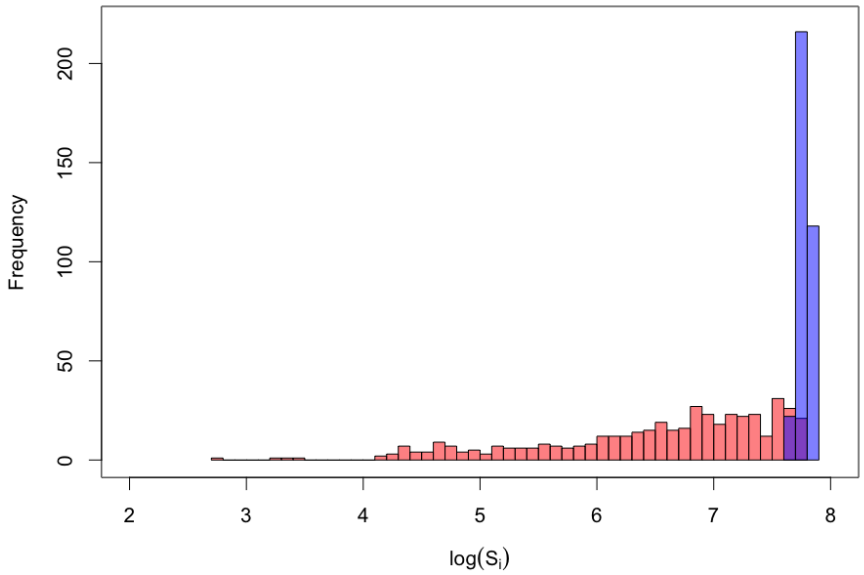


Figure 2: Histogram of the log-transformed survival times: observed (red) and censored (blue) observations.

j	1	2	3	4	5	6	7	8	9	10	MEAN
p_j	0.46	0.48	0.52	0.44	0.51	0.59	0.58	0.46	0.58	0.73	0.50

Table 2: Percentage of men in the data, stratified for each gap time j, $j = 1, \ldots, J$.

- *age* [years] of the patient at each hospitalization. Empirical mean of *age* at entrance in the study is 75.77 (s.e. 10.78). The empirical means of *age* stratified by gender are 78.44 (s.e. 10.02) for women and 73.14 (s.e. 10.86) for men, respectively. Age at the end of each gap time is included as a time-varying covariate in the model, but we find that there is no real difference in the posterior inference reported in Section 4 if we include only *age* at the entrance in the study and therefore we use this latter variable.

- *group*: indicator variable which identifies the clinical classification of the patient according to criteria detailed in Mazzali et al. (2016). In particular, the patients are classified based on indicators proposed by the Agency for Healthcare Research and Quality (AHRQ Quality Indicators, 2015) and HF codes as identified by the Center for Medicare and Medicaid Services (Evans et al., 2011). As a result, four different groups were defined (see Figure 2 and Table 2 in Mazzali et al., 2016): G1 denotes the group of patients having HF as the cause of admission or complicating another cardiac disease. G2 includes patients with Myocardial or cardiopulmonary diseases. G3 refers to patients with Acute HF as a complication of other diseases or for whom HF is reported as comorbidity. Finally, G4 identifies the remaining subjects (only three). The "group" variable is defined at the time of the first heart failure event, independently of subsequent events. As G1 represents the most frequent classification, as well as the most traditional characterization of hearth failure, we reduce the variable *group* to a binary covariate, which is set equal to 0 if the label of the patient is G1 (560 patients), and 1 otherwise (250 patients). Hence, *group* denotes the indicator of non-standard pathology.

- *rehab*: binary variable indicating if any time during the hospitalization is spent in a rehabilitation unit: 11.78% of the hospitalizations are spent partially or completely in a rehabilitation unit, corresponding to 29.01% of the patients.

- *ic*: binary variable indicating if at least a part of the hospitalization is spent in a intensive care unit. This happens in 11.95% of the hospitalizations, corresponding to 31.48% of patients.

- *n_com*: total number of comorbidities for each hospitalization. Table 3 reports the average number of comorbidities for all patients observed at the j-th gap times, for $j = 1, \ldots, J$.

j	1	2	3	4	5	6	7	8	9	10	MEAN
n_com	2.19	2.69	3.14	3.52	3.88	4.12	4.28	4.58	5.00	5.09	3.05

Table 3: Average number of comorbidities for hospitalization, stratified for each gap time j, $j = 1, \ldots, J$.

- *n_pro*: total number of surgical procedures for each hospitalization. Table 4 reports the average number of surgical procedures for all patients observed at the j-th gap times, for $j = 1, \ldots, J$.

j	1	2	3	4	5	6	7	8	9	10	MEAN
n_pro	0.10	0.15	0.10	0.12	0.07	0.08	0.08	0.05	0.03	0	0.11

Table 4: Average number of surgical procedures for hospitalization, stratified for each gap time j, $j = 1, \ldots, J$.

Note that 426 (52.59%) patients spend some time in either a rehabilitation or intensive care unit; none of the patients is admitted in both rehabilitation and intensive care unit at the same hospitalization; 129 (15.93%) patients enter a rehabilitation unit in at least one hospitalization, but never an intensive care one, while for 149 (18.40%) patients the opposite occurs.

Moreover, it is important to highlight that each gap time is calculated as the difference between two successive hospitalizations and, as such, it captures both the length of stay in hospital and the time between the discharge and the next hospitalization of the patient. In the analysis we include the value of the time dependent covariates measured at the end of each gap time, for example, $\boldsymbol{x}_{i1}$ refers to the covariates of patient i at the end of the first waiting time.

When fitting model (3)-(6), variables *gender*, *age* (at the first hospitalization) and *group* are treated as fixed covariates ($p = 3$), whereas *rehab*, *ic*, *n_com* and *n_pro* are time varying ($q = 4$). Moreover, *age*, *n_com* and *n_pro* have been standardized to have mean zero and variance one. The remaining covariates are binary. Therefore, the linear regression term in (3) for patient i at time j is given by.

$$\beta_{01}x_{i1} + \beta_{02}x_{i2} + \beta_{03}x_{i3} + \beta_{j1}x_{ij1} + \beta_{j2}x_{ij2} + \beta_{j3}x_{ij3} + \beta_{j4}x_{ij4},$$

where $\boldsymbol{x}_{ij} := (x_{i1}, x_{i2}, x_{i3}, x_{ij1}, x_{ij2}, x_{ij3}, x_{ij4})$; x_{i1} and x_{i2} correspond to indicators for gender (= 0 if the patient is a male) and pathology group (= 0 for standard pathology

G1), respectively, x_{i3} is the standardized age at the beginning of the study; x_{ij1} and x_{ij2} denote the binary variables *rehab* and *ic* for patient i at the j-th time, respectively, while x_{ij3} and x_{ij4} are the standardized number of comorbidities and number of surgical procedures of patient i during the j-th gap time. We assume that $\boldsymbol{z}_i = (x_{i1}, x_{i2}, x_{i3})$, i.e., the fixed covariates included in the survival regression component are gender, pathology and standardized age at the beginning of the study.

4 Posterior inference and cluster estimates

Posterior inference for the proposed model can be performed through a standard Gibbs sampler algorithm, which has been implemented in JAGS (Plummer, 2003) and run within the R software (R Core Team, 2015), through the R package `rjags`. We have run the MCMC sampler for $70,000$ iterations, discarding the first $10,000$ as burn-in and thinning every 12 iterations; the final sample size is $5,000$. We check through standard diagnostics criteria, such as those available in the R package CODA (Plummer et al., 2006), that convergence of the chain is satisfactory for most of the parameters.

In the analysis we select hyperparameter values that reflect lack of information, i.e., we opt for non-informative prior distributions. We choose $\eta_i^2 = \eta^2 \sim \text{inv-Gamma}(a_\eta, b_\eta)$ and the two-dimensional nonparametric prior component as in (7) with

$$\beta_0^2 = 100; a_\tau = 2.01; b_\tau = 1.01; \gamma_0^2 = 100; \psi_0^2 = 100$$
$$a_M = 0.3; b_M = 5; \alpha_0^2 = 100; a_\sigma = a_\eta = 2.01; b_\sigma = b_\eta = 1.01.$$

In particular, the prior choice for the variance parameters σ_i^2, η^2 implies a priori marginal expected value of 1 and an a priori variance equal to 100 for both of them. Analogously, we specify a large value for the standard deviation of the regression coefficients $\boldsymbol{\beta}, \boldsymbol{\gamma}$ and ψ. We opt for a uniform distribution between 0.3 and 5.0 as prior for the total mass M of the Dirichlet Process. This choice corresponds to a prior belief that the expected number of clusters is large, i.e., $\mathbb{E}(K) = 15.33$. The lower bound of the support of the marginal prior for M, 0.3, is specified to avoid computational difficulties in JAGS caused by small weights in the stick-breaking representation of the DP (see Ohlssen et al., 2007, for instance). MCMC implementation of the DP in JAGS requires a truncation of the process, i.e., truncating the stick-breaking representation (1) to a prefixed level C. This strategy, on which many MCMC schemes for the DP are based, obviously leads to a truncation error, which depends on the choice of C. As

suggested in Ohlssen et al. (2007) we fix C equal to 40, so that, in case M assumes values close to the upper bound of its prior support, the truncation error is negligible.

4.1 Posterior inference

We now present the inference results for the regression parameters in order to understand how covariates influence the recurrent events distribution and survival, regardless of the underlying structure of the trajectories (which is captured by the subject-specific parameters).

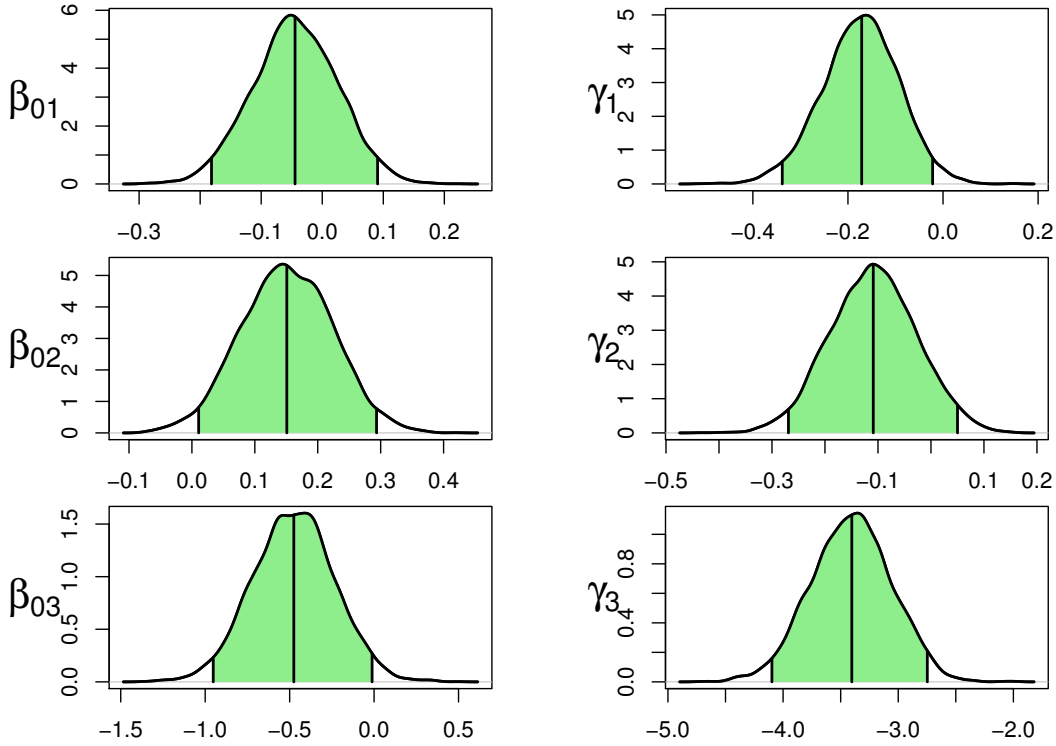


Figure 3: Posterior marginal densities of the regression coefficients $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}$ of the time-homogeneous covariates. The shaded regions correspond to 95% credible intervals. Top row corresponds to the effect of *gender* on gap (left) and survival (right) times, middle row of *group*, bottom row of *age*.

Figure 3 shows the 95% credible intervals for the posterior marginals of the fixed effect regression parameters $(\beta_{01}, \beta_{02}, \beta_{03})$ and $(\gamma_1, \gamma_2, \gamma_3)$. In what follows we assume that a covariate has an effect on the response if its 95% credible interval does not cover zero. As such, we observe that:

- there is no evident effect of *gender* on the gap times, but an effect is detectable

on the survival times (with a negative effect on the survival time for women), see the top plots in Figure 3. Sample average survival times by gender confirm that women have slightly shorter survival times (7.07 in the log scale, with standard deviation equal to 0.96; 7.11 and 0.96 are the corresponding sample average and standard deviation for men);

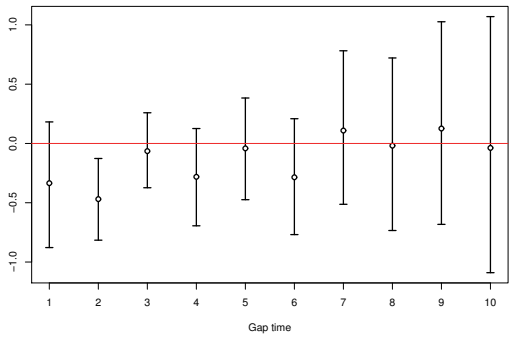
- the *group* variable seems to be relevant for the gap times (patients with non-standard pathology, i.e. not in group G1, have larger gap times); however it does not influence the survival time;

- the *age* variable is a predictor of both gap times and survival (see the bottom plots in Figure 3): in particular, there is evidence that the average time between hospitalizations is shorter for older patients.

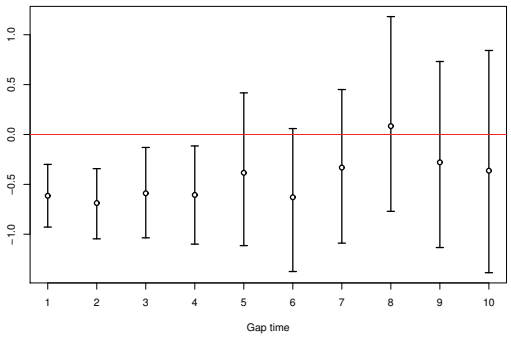
Time-varying covariates are reported at the end of each gap time. We need to take this into account when analysing the results for the regression coefficients of such covariates. In general time-varying covariates do not have a strong effect on the recurrence process, except for few early gap times and, as expected, the uncertainty on the effect estimates increases over time, due to the smaller number of available observations.

Figure 4 displays the 95% CIs of the posterior marginal densities of the regression coefficients $(\beta_{j1}, \beta_{j2}, \beta_{j3}, \beta_{j4})$ of the time-varying covariates. It is evident that *rehab* and *ic* have an effect on the distribution of waiting times, as patients with *rehab* or *ic* equal 1 show shorter gap times (the effect of *ic* is stronger than *rehab*). Moreover, the number of comorbidities does not seem to be influential and it appears that a large number of surgical procedures yields frequent hospitalizations at the beginning of the study, followed by an opposite effect (delayed hospitalizations) for later gap times.

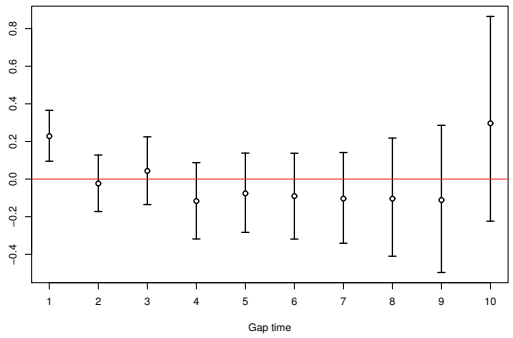
Figure 5a shows the posterior predictive density for a hypothetical new patient of the random effect parameters $\alpha^\star$ which is modelled using Dirichlet Process prior jointly with σ^2. This distribution is multimodal, indicating a clustering structure among patients. Moreover, in Figure 5b we plot the marginal posterior distribution of the parameter ψ, which links the survival outcome to the recurrent event process and reflects the strength of the relationship between the two processes. In our application, this distribution is centred away from zero, on the positive axis, indicating that as the time between hospitalizations widens, the probability of survival increases as well.



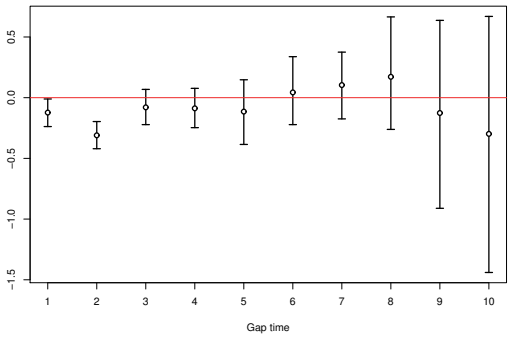
(a) Rehabilitation status.



(b) Intensive care unit status.

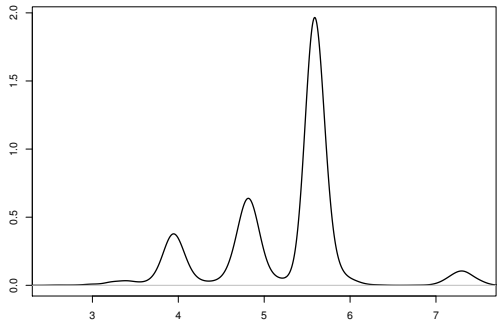


(c) Number of comorbidities.

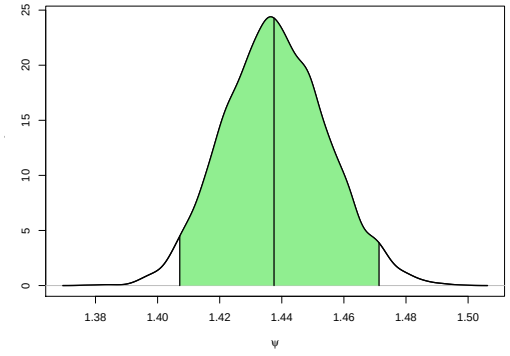


(d) Number of surgical procedures.

Figure 4: 95% credible intervals of posterior marginal densities of the regression coefficients of the time-varying covariates. The red line corresponds to 0.



(a) Posterior predictive density of $\alpha^\star$.



(b) Posterior marginal density of ψ.

Figure 5: Posterior predictive density of $\alpha^\star$ and posterior marginal density of ψ.

4.2 Clustering

As mentioned in the Introduction, our model, described by (3)-(4) and (6)-(7), induces a prior on the partition of the subjects in the sample (Barcella et al., 2015).

We denote with $\boldsymbol{\rho}$ a random partition of the patients. In this framework it is then straightforward to perform posterior inference on the clustering structure and to obtain a posterior estimate of the parameter $\boldsymbol{\rho}$ from the MCMC output. Here we report the clustering allocation $\boldsymbol{\rho}$ that minimises the posterior expectation of Binder's loss function (Binder, 1978) under equal misclassification costs, a common choice in the applied Bayesian nonparametric literature (Lau and Green, 2007); see Argiento et al. (2014) for computational details. Note that the parameters (α_i, σ_i^2) determine the clustering of patients. Since the parameter α_i links the failure and recurrence processes, our modelling strategy allows the clustering to depend on both gap times trajectories and survival outcome. The estimated partition contains 6 clusters. In Table 5 we report cluster-specific summary statistics.

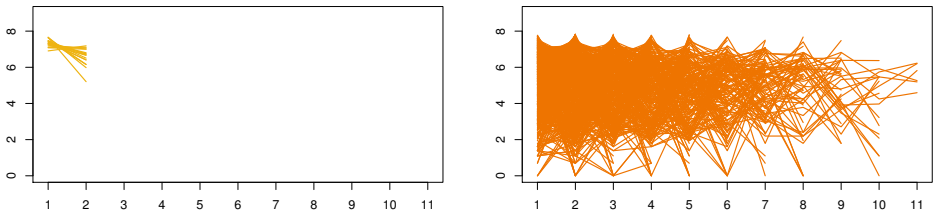
Cluster	Size	$\overline{u}_n$	$\overline{y}_n$	$\overline{age}$	$\bar{n}_i$	% censored
1	17 (2.16 %)	7.769 (0.048)	7.342 (0.209)	74.706 (9.980)	1	100
2	537 (68.15 %)	7.561 (0.381)	4.883 (1.696)	75.973 (10.592)	3.669	61.82
3	14 (1.78 %)	4.742 (0.651)	3.576 (0.632)	73.214 (10.245)	3	0
4	104 (13.2 %)	6.643 (0.570)	4.568 (1.204)	75.794 (10.867)	4.356	1.92
5	112 (14.21 %)	5.546 (0.818)	3.864 (1.366)	75.157(11.764)	2.982	0.89
6	4 (0.51 %)	7.764 (0.054)	5.753 (0.557)	79 (13.441)	6	100

Table 5: Cluster specific sample summary statistics: size, average survival time (standard deviation), average gap time (standard deviation), average number of gap times per trajectory, censoring rate

Trajectories of the gap times, $Y_{ij}, j = 1, \ldots, n_i$, for all the patients are displayed in Figure 6 for each cluster. Moreover, in Figure 7 we show observed and censored survival times and corresponding kernel density estimates for the six clusters above.

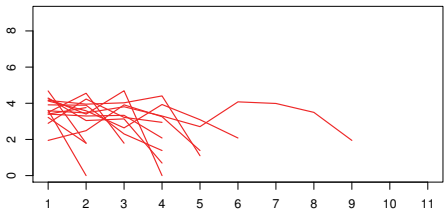
The largest cluster of patients (68.15% of the patients), denoted as cluster 2 in Table 5, is characterized by large survival times and long gap-time trajectories. Cluster 1 and 6 include only censored observations and are characterized by large survival times, but also by larger gap times compared to Cluster 2 and 4. Clusters 3, 4 and 5 present shorter time intervals between hospitalizations compared to the others clusters as well as shorter survival times (see Table 5 for details). The percentage of patients with standard pathology (i.e. *group*=0) is similar to the overall rate ($\simeq 70\%$) in each cluster but in Cluster 1, where it is 53%.

Finally, Figure 8 shows the posterior distribution of K, the number of clusters in

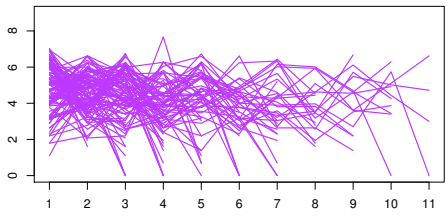


(a) Cluster 1

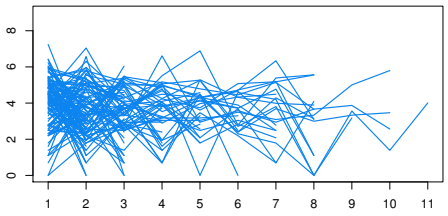
(b) Cluster 2



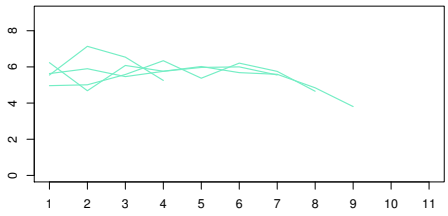
(c) Cluster 3



(d) Cluster 4



(e) Cluster 5



(f) Cluster 6

Figure 6: Trajectories of gap times. Lines connect the gap times for each observation.

the sample, whose mode is at 6.

5 Goodness-of-fit and model selection

In Section 4 of the paper we fit our model when the variance of the survival time is assumed to be the same for all the subjects. We also consider the model in which we introduce a subject specific variance, η_i, as in (4). We refer to the first models as DP_2 and to the second one as DP_3, respectively, since the prior component for the subject specific effects is a DP defined on $\mathbb{R}^2$ (see (7)) or $\mathbb{R}^3$ (see (5)).

We compare the models using two measures of predictive performance: the Watanabe-Akaike information criterion (WAIC) and the Brier score. WAIC is a fully Bayesian approach for estimating the predictive accuracy of the dataset: it is obtained computing the log point-wise posterior predictive density and then adding a correction (a penalty) for the effective number of parameters to penalise overfitting. In particular,

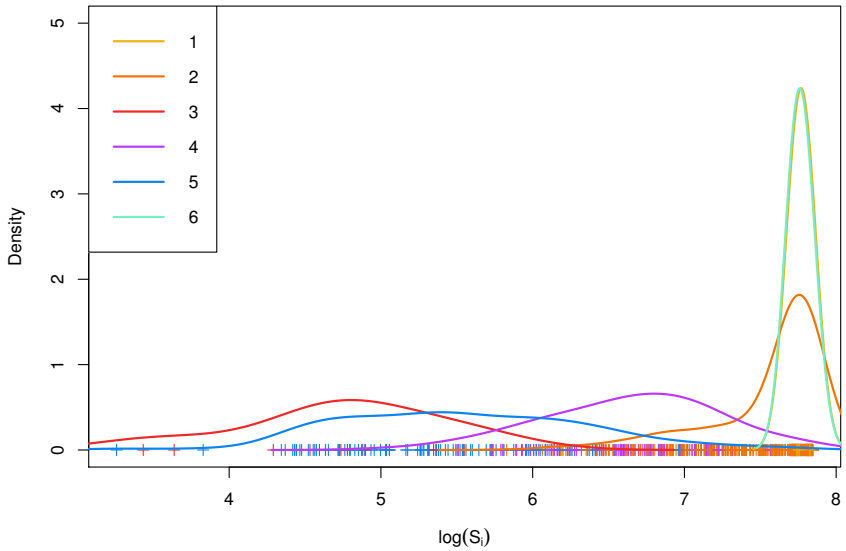


Figure 7: Survival times coloured according to cluster allocations. Solid lines correspond to kernel density estimates.

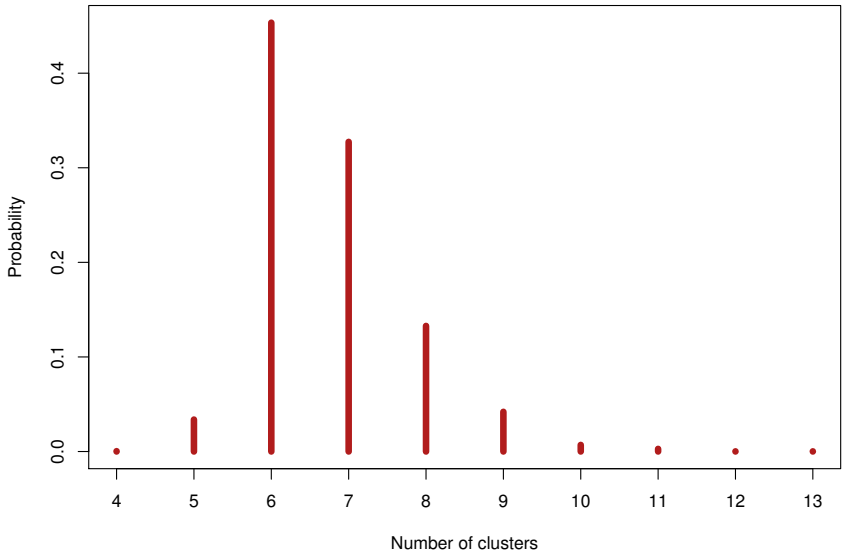


Figure 8: Posterior distribution of the number of clusters.

$WAIC$ is defined as $lppd - p_{WAIC1}$, where $lppd$ is the log pointwise predictive density, i.e., the product (in the log scale) of the conditional densities (evaluated at y_{ij}), of Y_{ij}, given all the data, and then adding the bias correction p_{WAIC1}, which is similar to the bias correction in the definition of the DIC (see Gelman et al., 2014, for details). In our case, the WAIC values computed for the two models (see Table 6) indicate a better

performance of DP_3.

	DP_2	DP_3
WAIC	-5979.591	-5973.26

Table 6: WAIC values for the model DP_2 and DP_3; the penalization used is p_{WAIC1} as in Gelman et al. (2014).

The second criterion is the Brier score (Brier, 1950), that measures the accuracy of in-sample prediction. As the Brier score is usually evaluated for binary classification problems, we need to dichotomise our prediction to adapt the Brier score to continuous data, as proposed by Barcella et al. (2016). Therefore, we are interested in predicting whether the survival time of an individual is above or below a specific threshold. We use the quartiles of the observed survival times as thresholds. Therefore, we discretise the observed data so that $\tilde{u}_i^{(k)} = 0$ if $\tilde{u}_i \leq Q_k$ and $\tilde{u}_i^{(k)} = 1$ if $\tilde{u}_i > Q_k$, where Q_k is the k-th quartile of the data, $k = 1, 2, 3$. At each iteration of the Markov chain, for each patient i we evaluate the predictive probability f_i of obtaining a survival time larger than the specified threshold and we then compare it with the observed value. The Brier Score is defined as follows

$$\mathrm{BR}_k = \frac{1}{N} \sum_{i=1}^{N} \left(f_i^{(k)} - \tilde{u}_i^{(k)} \right)^2,$$

that is, for each patient we compute the difference between the predictive probability of the observation to be above the threshold Q_k and the observed value $\tilde{u}_i^{(k)}$. Small values of the Brier statistic indicates good classification performance.

Figure 9 displays the boxplots of the Brier Score under 6 scenarios (model DP_2 and DP_3 for each of the three quartiles Q_1, Q_2 and Q_3).

As expected, the two models perform better when estimating the probability that the prediction and the true observation lie at the same side of the first quartile. Indeed, the time-to-event data in this application consist of strongly left skewed observations. Therefore, it is more difficult to predict large survival times. In general, however, we observe that for each model and for each quartile, the Brier Score is smaller than the threshold of 25%, which represents the case in which the prediction of interest is equivalent to a coin toss. Moreover, model DP_2 outperforms DP_3 for the first quartile, but DP_3 is more accurate for the prediction of higher survival times.

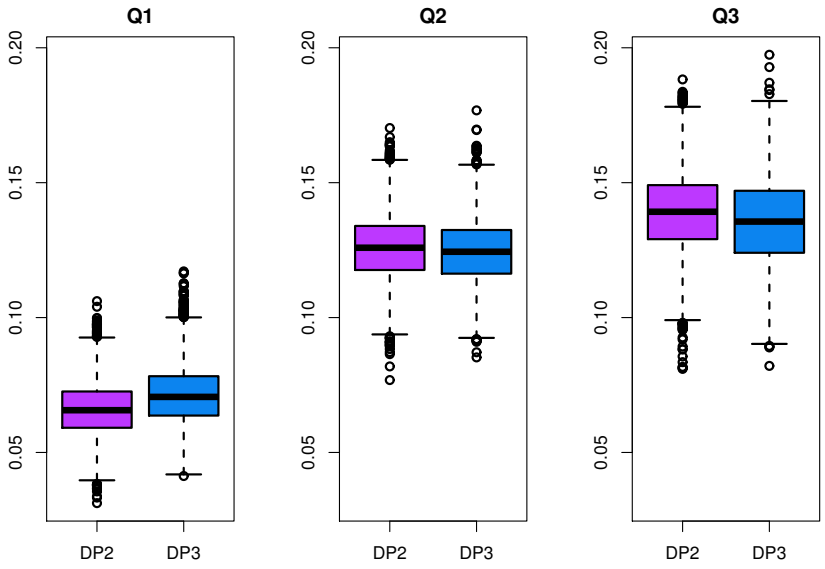


Figure 9: Boxplots of the Brier scores corresponding to the three quartiles of the data for model DP_2 (purple) and DP_3 (blue).

6 Out-of-sample predictive inference

In order to asses the predictive performance of our model, 22 patients (5% of the total sample size) are randomly selected and removed from the training set, leaving 788 patients in the sample. The training set includes patients with at most nine observed gap times. We fit the model (see details in Section 4 of the paper) to the reduced dataset, and we predict the distribution of the gap time as well as of survival time for the excluded patients using their covariate values. We then compare our predictions with the observed data. To summarise our results, we compute the Mean Squared Error between predicted and observed waiting times, i.e., $\text{MSE}_j = \sum_{i=1}^{N_j^*}(y_{ij}^* - y_{ij})^2$, where N_j^* is the number of patients experiencing at least j recurrences in the test set, and y_{ij}^* denotes the median of the predictive density for patient i at gap time j in the test set, whose corresponding observed value is y_{ij}; the values obtained for all 22 patients, at each time j and overall, are shown in Table 7 for the two models discussed in Section 5. No significant differences in predictive performance between the two models are noticeable.

Figure 10 displays the predicted trajectories of gap times for a subset of 9 randomly selected patients. In our model, we assume that each patient is characterised by a random effect influencing the mean of every gap time and a covariate effect. Therefore,

$\mathbf{MSE}_j$	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	$j=8$	$j=9$	$\sum_{j=1}^{9} \text{MSE}_j$
DP_2	2.05	2.40	1.13	3.71	2.24	1.73	0.33	3.44	1.66	**18.70**
DP_3	2.06	2.43	1.16	3.69	2.21	1.74	0.33	3.41	1.74	**18.77**

Table 7: Mean squared error of the out-of-sample prediction for each gap time under models DP_2 and DP_3.

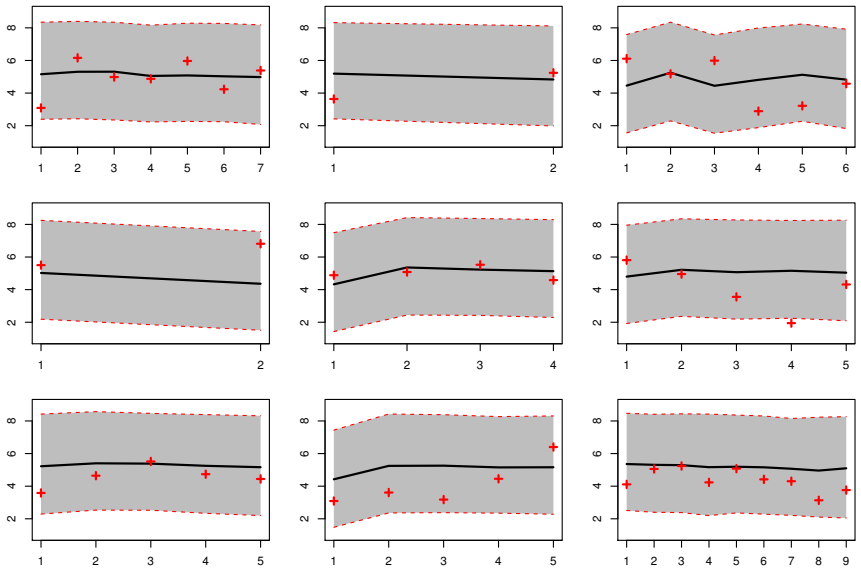


Figure 10: Predictive distribution of the gap time trajectory for 9 randomly selected patients. The solid black line represents estimated posterior medians of the waiting time trajectory, the shadowed area is the 95% credible interval. Red crosses represent observed values.

in this model the oscillations around the mean value of the gap times are given by a change in the values of the covariates. Finally, in addition to the Brier score presented in Section 5, we evaluate the 95% predictive credible interval for the the survival time of each of the 22 patients in the test set to assess the ability of the model to predict the survival component. Under both models, 21 out of 22 observed survival times are included in the 95% CIs.

7 Comparison with other models

We now compare our results to those obtained employing one of the most popular semi-parametric models in survival analysis with covariates: the Cox proportional hazards (PH) model. In this model, the hazard function for the survival time is specified as

$$h_i(t|\boldsymbol{z}_i, \boldsymbol{\gamma}) = h_0(t)e^{\boldsymbol{z}_i^T\boldsymbol{\theta}}, \tag{8}$$

where $h_0(\cdot)$ is the baseline hazard rate, $\boldsymbol{z}_i^T$ is a vector of covariates and $\boldsymbol{\theta}$ is a vector of regression coefficients. Under model (8), the larger $\boldsymbol{z}_i^T\boldsymbol{\theta}$, the larger the hazard of the event. To make the comparison with the model described in Section 2 as fair as possible, we include in $\boldsymbol{z}_i$ the same fixed covariates (z_{i1}, z_{i2}, z_{i3}) influencing the time S_i (i.e. gender, group and age), but we also add a fourth covariate z_{i4} representing the mean waiting times (in log-scale) of each individual. This latter covariate is supposed to give a heuristic approximation of the random effect α_i in the proposed Bayesian semiparametric model.

The application of the PH model to the HF data show consistent results with those obtained by our model. In particular, all the covariates are significant, and the Hazard Ratios are:

- $e^{\theta_1} = 1.42$ for gender, indicating that female patients have an hazard function which is greater than male patients, all the other covariates being equal;
- $e^{\theta_2} = 141.38$ for age at the first event, which corresponds to strong evidence that age is the main driver of the risk of death;
- $e^{\theta_3} = 1.32$ for group, implying that patients with non-standard HF are at higher risk;
- $e^{\theta_4} = 0.68$, which indicates that longer gap times can result in longer survival times.

Thus, the effect of gender, group and age is the same as in the proposed model. In general, the uncertainty associated to the estimates obtained under our model is higher as we also account for uncertainty in the gap times distribution. Nevertheless, our approach introduces extra flexibility when estimating the distribution of the survival times, as it allows the identification of different risk groups.

For a fairer comparison, we also fit to our data the joint frailty model developed by Rondeau et al. (2007) and implemented in the R package `frailtypack` (Rondeau et al.,

2012). Using this approach it is possible to estimate jointly the two hazard functions associated with recurrent and terminal events. In this context, the dependence is modelled through a common frailty term, v_i, that takes into account the heterogeneity in the data. The frailty term affects differently the two hazards functions. Conditional on the frailty term v_i, the joint frailty model is given by:

$$\begin{cases} r_{ij}(t|v_i) = v_i r_0(t) \exp\left\{\boldsymbol{x}_{ij}^T \boldsymbol{\beta}\right\} & \text{(recurrent events)} \\ \lambda_i(t|v_i) = v_i^{\xi} \lambda_0(t) \exp\left\{\boldsymbol{z}_i^T \boldsymbol{\gamma}\right\} & \text{(terminal event)} \end{cases}$$

where $r_0(t)$ (resp. $\lambda_0(t)$) is the recurrent (resp. terminal) event baseline hazard function, $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_i$ are vector of covariates for individual i at time j, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the corresponding vectors of regression parameters and v_i are *iid* Gamma random variables with mean 1 and variance κ. If the parameter ξ is equal to zero, then the dependence between the two processes is captured only by specifying common covariates.

In fitting this model to the HF data we use the same covariates as described in Section 3. Notice that the model allows including time-varying covariates in the recurrent event process, but, differently from our approach, the effect is assumed constant over time.

We report the results of the joint frailty model in Table 8. The application of this model to the HF data yields conclusions that are in general consistent with the ones obtained with the Bayesian semiparametric model. The coefficients for *ic* and *npro* are significant. This can be explained by the fact that in this case the effect is assumed to be constant and therefore the model can borrow strength across gap-times. The most notable difference is the coefficient of gender in the survival regression (although the p-value is 0.04) that has an opposite effect in the joint frailty model compared to our approach. Nevertheless, our results are in agreement with both the Kaplan-Meier estimator stratified by gender (results not shown) and with the Cox Proportional Hazard analysis. Finally, the estimate of κ is 0.2505 (s.e. 0.023), which implies that there is heterogeneity between subject not explained by the covariates, as confirmed by the clustering structure obtained by our model. The estimate of ξ is 4.70 (s.e. 0.3461); this means that the incidence of recurrences is positively associated with death, once again in agreement with our model.

Recurrences:			
	Coefficient	Std. Dev.	p-value
gender	-0.0709	0.0503	0.15
group	0.0375	0.0542	0.49
age	0.8027	0.1676	$1.68 \cdot 10^{-6}$
rehab (TV)	-0.01	0.0633	0.87
ic (TV)	0.4162	0.0667	$4.36 \cdot 10^{-10}$
ncom (TV)	0.007	0.0238	0.78
npro (TV)	0.1194	0.0220	$6.45 \cdot 10^{-8}$
Terminal event:			
	Coefficient	Std. Dev.	p-value
gender	-0.3206	0.1669	0.04
group	-0.1970	0.1814	0.27
age	7.4208	0.6478	0

Table 8: Estimated regression coefficients of the joint frailty model, as implemented in the R package `frailtypack`. TV denotes a time varying covariate.

8 Discussion

Treating and managing appropriately Heart Failure (HF) patients is a major public health issue. Indeed, HF is one of the major causes of hospitalization and death in adult population and the main reason for hospital admission in patients aged over 65 years in western countries; see Desai and Stevenson (2012). As the population ages and the prevalence of heart failure increases, expenditures related to the care of these patients are climbing dramatically. As a result, the health care industry must develop strategies to contain the economic burden without compromising the effectiveness of the care. It has become evident that there is an urgent need for methods supporting healthcare management by improving evidence-based practice. To this aim, it is essential to gain a deeper understanding of the clinical factors contributing to lengthen/decrease short-, middle- and long-term survival jointly with the length and recurrence of hospitalizations. However, most analysis focus only on the primary clinical outcomes (such as death or time to first re-hospitalization), ignoring the information contained in the recurrent event process and the relationship between multiple hospital admissions and

patient's survival.

In this paper we propose a joint semiparametric model for recurrent hospitalizations due to HF and time to death. Our approach jointly models survival and the hospitalizations times, specifying a DP as random effect distribution of the frailty parameter that links the survival and gap time trajectories. This strategy allows us to introduce extra flexibility in the model to account for patients heterogeneity and identify different risk groups. Other Bayesian nonparametric priors could be employed, at the cost of more expensive computations. An important feature of the model is to be able to take into account the dependent censoring of gap times by death since in many applications, such as ours, there is a strong relationship between event recurrences and termination. Time homogeneous and time-varying covariates are easily incorporated within a regression framework. Specification of different survival and/or gap time distributions can also be easily accommodated. This approach can be extended to include time dependency between gap times through the inclusion of, for example, autoregressive terms. Main advantages of the proposed methodology are wide applicability, ease of interpretation and efficient computations.

Our results show that women tend to have lower survival times. This can be due to the fact women are much older than men when entering the study and they are more likely to die because the comorbidity load increases with age. The effect of the variable *group* on gap times reflect different protocols of HF treatments. The strong effect of *age* on both recurrent and survival process is not surprising as older age is usually associated with worse health conditions and comorbidity load. Furthermore, we have investigated the effect over time of the time-varying covariates, highlighting possible temporal patterns. For example, it is evident that admission to an intensive care unit shortens early gap times, as well as rehabilitation. This can be explained by the fact that usually these variables are associated with more serious complications. Estimates of the effect on later gap times are associated with wider credible interval, due to less observations available. Moreover, the model is able to account for patient-specific heterogeneity through the data-driven clustering of patients based on their re-hospitalizations trajectory and survival outcome. In future work we will extend the methodology to a much richer dataset, which will include a wider patient population and new potential explanatory variables. Moreover, an hospital effect and spatial information can be easily incorporated in the model, as well as variable selection strategies. These future directions of research will most likely require generalising

the methodology to combine aggregated and individual level information. Finally, the model assumes a time-homogeneous effect of the gap times on survival, which can be a strong modelling assumption when many subsequent hospitalizations are associated with a higher risk of death. This limitation can be overcome by, for example, specifying an autoregressive model for the random effect parameter α_i (Tallarita et al., 2016).

References

AHRQ Quality Indicators (2015). "Heart Failure Mortality Rate, Technical Specifications, version 5.0, March 2015." `http://www.qualityindicators.ahrq.gov/Downloads/Modules/IQI/V50/TechSpecs/IQI_16_Heart_Failure_Mortality_Rate.pdf`.

Argiento, R., Cremaschi, A., and Guglielmi, A. (2014). "A density-based algorithm for cluster analysis using species sampling Gaussian mixture models." *Journal of Computational and Graphical Statistics*, 23, 1126–1142.

Barcella, W., De Iorio, M., and Baio, G. (2015). "A comparative review of variable selection techniques for covariate dependent Dirichet process mixture models." *arXiv preprint arXiv:1508.00129*.

Barcella, W., Iorio, M. D., Baio, G., and Malone-Lee, J. (2016). "Variable selection in covariate dependent random partition models: an application to urinary tract infection." *Statistics in Medicine*, 35, 1373–1389.

Binder, D. A. (1978). "Bayesian cluster analysis." *Biometrika*, 65, 31–38.

Brier, G. W. (1950). "Verification of forecasts expressed in terms of probability." *Monthly Weather Review*, 78, 1–3.

Brown, E. and Ibrahim, J. (2003). "A Bayesian semiparametric joint hierarchical model for longitudinal and survival data." *Biometrics*, 59, 221–228.

Conlon, A., Taylor, J., and Sargent, D. J. (2014). "Multi-state models for colon cancer recurrence and death with a cured fraction." *Statistics in Medicine*, 33, 1750–1766.

Cook, R. J. and Lawless, J. F. (2007). *The statistical analysis of recurrent events*. Springer, New York.

Desai, A. S. and Stevenson, L. W. (2012). "Rehospitalization for heart failure." *Circulation*, 126, 501–506.

Evans, M. A., Pope, G. C., Kautter, J., Ingber, M. J., Freeman, S., Sekar, R., and Newhart, C. (2011). "Evaluation of the CMS-HCC Risk Adjustment Model." *CfMM Services, Editor*. `https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/downloads/evaluation_risk_adj_model_2011.pdf`.

Ferguson, T. S. (1973). "A Bayesian analysis of some nonparametric problems." *The Annals of Statistics*, 1, 209–230.

Gelman, A., Hwang, J., and Vehtari, A. (2014). "Understanding predictive information criteria for Bayesian models." *Statistics and Computing*, 24, 997–1016.

Huang, X. and Liu, L. (2007). "A joint frailty model for survival and gap times between recurrent events." *Biometrics*, 63, 389–397.

Lau, J. W. and Green, P. J. (2007). "Bayesian model-based clustering procedures." *Journal of Computational and Graphical Statistics*, 16, 526–558.

Mazzali, C., Maistriello, M., Ieva, F., and Barbieri, P. (2015). "Methodological issues in the use of administrative databases to study heart failure." In *Advances in Complex Data Modeling and Computational Methods in Statistics*, 149–160. Springer.

Mazzali, C., Paganoni, A. M., Ieva, F., Masella, C., Maistrello, M., Agostoni, O., Scalvini, S., and Frigerio, M. (2016). "Methodological issues on the use of administrative data in healthcare research: the case of heart failure hospitalizations in Lombardy region, 2000 to 2012." *BMC Health Services Research*, 16, 234.

Ohlssen, D. I., Sharples, L. D., and Spiegelhalter, D. J. (2007). "Flexible random-effects models using Bayesian semi-parametric models: applications to institutional comparisons." *Statistics in Medicine*, 26, 2088–2112.

Ouyang, B., Sinha, D., Slate, E. H., and Van Bakel, A. B. (2013). "Bayesian analysis of recurrent event with dependent termination: an application to a heart transplant study." *Statistics in Medicine*, 32, 2629–2642.

Plummer, M. (2003). "JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling." In *Proceedings of the 3rd international workshop on distributed statistical computing, Vienna*.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). "CODA: Convergence Diagnosis and Output Analysis for MCMC." *R News*, 6, 7–11.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rockova, V., Lesaffre, E., Luime, J., and Löwenberg, B. (2012). "Hierarchical Bayesian formulations for selecting variables in regression models." *Statistics in Medicine*, 31, 1221–1237.

Rondeau, V., Mathoulin-Pelissier, S., Jacqmin-Gadda, H., Brouste, V., and Soubeyran, P. (2007). "Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events." *Biostatistics*, 8, 708–721.

Rondeau, V., Mazroui, Y., and Gonzalez, J. R. (2012). "frailtypack: an R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation." *J Stat Softw*, 47, 1–28.

Schmoor, C., Schumacher, M., Finke, J., and Beyersmann, J. (2013). "Competing risks and multistate models." *Clinical Cancer Research*, 19, 12–21.

Sethuraman, J. (1994). "A constructive definition of Dirichet priors." *Statistica Sinica*, 4, 639–650.

Sinha, D., Maiti, T., Ibrahim, J. G., and Ouyang, B. (2008). "Current methods for recurrent events data with dependent termination." *Journal of the American Statistical Association*, 103, 866–878.

Tallarita, M., De Iorio, M., Guglielmi, A., and Malone-Lee, J. (2016). "Bayesian Nonparametric Modelling of Joint Gap Time Distributions for Recurrent Event Data." *arXiv preprint arXiv:1607.08141*.

Yu, Z. and Liu, L. (2011). "A joint model of recurrent events and a terminal event with a nonparametric covariate function." *Statistics in Medicine*, 30, 2683–2695.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

23/2017 Quarteroni, A.; Vergara, C.
Computational models for hemodynamics

24/2017 Domanin, M.; Bissacco, D.; Le Van, D.; Vergara, C.
Computational fluid-dynamic comparison between patch-based and direct suture closure techniques after carotid endarterectomy

21/2017 Talska, R.; Menafoglio, A.; Machalova, J.; Hron, K.; Fiserova, E.
Compositional regression with functional response

22/2017 Bartezzaghi, A.; Dede', L.; Quarteroni, A.
Biomembrane modeling with Isogeometric Analysis

20/2017 Albrecht G.; Caliò F.; Miglio E.
Fair surface reconstruction through rational triangular cubic Bézier patches

19/2017 Giovanardi, B.; Formaggia, L.; Scotti, A.; Zunino P.
Unfitted FEM for modelling the interaction of multiple fractures in a poroelastic medium

18/2017 Ambartsumyan, I.; Khattatov, E.; Yotov, I.; Zunino, P.
A Lagrange multiplier method for a Stokes-Biot fluid-poroelastic structure interaction model

14/2017 Bruggi, M.; Parolini, N.; Regazzoni, F.; Verani, M.
Finite Element approximation of an evolutionary Topology Optimization problem

16/2017 Ghiglietti, A.; Scarale, M.G.; Miceli, R.; Ieva, F.; Mariani, L.; Gavazzi, C.; Paganoni, A.M.; E
Urn models for response-adaptive randomized designs: a simulation study based on a non-adaptive randomized trial

15/2017 Tagliabue, A; Dede', L.; Quarteroni A.
Complex blood flow patterns in an idealized left ventricle: a numerical study