MOX-Report No. 88/2023

Joint modelling of recurrent and terminal events with discretely-distributed non-parametric frailty: application on re-hospitalizations and death in heart failure patients

Masci, C.; Spreafico, M.; Ieva, F.

Joint modelling of recurrent and terminal events with discretely-distributed non-parametric frailty: application on re-hospitalizations and death in heart failure patients

Chiara Masci[1,†]
chiara.masci@polimi.it

Marta Spreafico[2,†]
m.spreafico@math.leidenuniv.nl

Francesca Ieva[1,3]
francesca.ieva@polimi.it

[1]MOX Lab, Department of Mathematics, Politecnico di Milano, Milan 20133, Italy
[2]Mathematical Institute, Leiden University, Leiden 2333 CA, The Netherlands
[3]Human Technopole, Health Data Science Center, Milan 20157, Italy

[†]*These authors contributed equally to this article.*

November 7, 2023

Abstract

In the context of clinical and biomedical studies, *joint frailty models* have been developed to study the joint temporal evolution of *recurrent and terminal events*, capturing both the heterogeneous susceptibility to experiencing a new episode and the dependence between the two processes. While discretely-distributed frailty is usually more exploitable by clinicians and healthcare providers, existing literature on joint frailty models predominantly assumes continuous distributions for the random effects. In this article, we present a novel *joint frailty model* that assumes *bivariate discretely-distributed non-parametric frailties*, with an unknown finite number of mass points. This approach facilitates the identification of latent structures among subjects, grouping them into sub-populations defined by a shared frailty value. We propose an estimation routine via Expectation-Maximization algorithm, which not only estimates the number of subgroups but also serves as an unsupervised classification tool. This work is motivated by a study of patients with Heart Failure (HF) receiving ACE inhibitors treatment in the Lombardia region of Italy. Recurrent events of interest are hospitalizations due to HF and terminal event is death for any cause.

Keywords: Recurrent events; Joint frailty models; Discrete frailty; Non parametric frailty; Heart Failure

1 Introduction

Recurrent or *repeated events* are common in many clinical and biomedical studies, as patients usually experience the same event multiple times. Typical situations are follow-up visits, hospital admissions, tumour relapses, heart attacks and many others. In the recurrent event framework, classic survival approaches are not suitable as they discard the correlation between subsequent events in the same subject. A wide literature about recurrent events modelling has hence flourished in past years (Kleinbaum and Klein, 1996; Therneau and Grambsch, 2000; Cook and

Lawless, 2007; Aalen *et al.*, 2008; Amorim and Cai, 2015; Ozga *et al.*, 2018). Among others, *frailty models* handle repeated episodes by introducing a random effect which takes a common value for each group of dependent observations (Hougaard, 1995, 2012; Kleinbaum and Klein, 1996; Therneau and Grambsch, 2000; Rondeau *et al.*, 2003, 2006; Cook and Lawless, 2007; Aalen *et al.*, 2008; Amorim and Cai, 2015; Gasperoni *et al.*, 2020). Given the existence of heterogeneous susceptibility to the risk of recurrent events among subjects, the random term can describe the excess risk or frailty of different individuals, accounting for unexplained heterogeneity not covered by observed covariates. Frailty models are particularly well suited to handle hierarchical data structures, such as subjects nested within groups (e.g., patients nested within hospitals). Nonetheless, an individual's recurrent process duration may be influenced by a terminal event, such as study end, loss to follow-up, or death. Death can prematurely end repeated events, and the terminal event time might be affected by the recurrent event history. Increased occurrences of serious events (e.g., re-hospitalizations) often raise the risk of death, challenging the assumption of independent censoring. As a result, joint frailty models of recurrent and terminal processes has gained significant attention.

Joint frailty models analyse both processes over time, treating the terminal process as informative censoring and accounting for their dependence. They capture both the correlation among repeated events and the dependence between repeated and terminal processes by incorporating random effects in both hazard functions. Lancaster and Intrator (1998) initially proposed a parametric model for repeated episodes via Poisson process with a rate function that shares the same subject-specific frailty as the time-to-death hazard, assuming the two processes independent given the frailty term. Huang and Wolfe (2002) introduced a joint frailty model for clustered data with informative censoring, sharing a log-normal frailty between censoring and failure rates at the cluster level. Liu *et al.* (2004), Huang and Liu (2007), and Rondeau *et al.* (2007) considered joint models with shared frailty terms that apply differently for the two hazard functions. Liu *et al.* (2004) and Huang and Liu (2007) used gamma frailty, focusing on time to events (i.e., calendar times) and time between events (i.e., gap times), respectively, and employing a Monte Carlo Expectation–Maximization algorithm for estimation. Alternatively, Rondeau *et al.* (2007) proposed a non-parametric penalized likelihood estimation method, accommodating gamma and log-normal shared frailty and handling both calendar and gap times. This approach also estimates (smoothed) hazard functions, which often have a meaningful interpretation in epidemiological studies. Zeng and Lin (2009) generalized joint shared-frailty models using a variety of transformation models, including various possible multivariate random-effects distributions. Further extensions included Bayesian non-parametric approaches (Paulon *et al.*, 2020), handling various situations like zero-inflated recurrent events (Liu *et al.*, 2016), nested clustered data for family studies (Choi *et al.*, 2020), cure fraction models (Tawiah *et al.*, 2020; Talebi-Ghane *et al.*, 2021), and non-proportional hazards through generalized survival models (Chauvet and Rondeau, 2023).

Moved by the need for a more flexible approach enabling two correlated random effects to jointly model the dependence between recurrent event and terminal event hazard rates, Ng *et al.* (2023) adapted the formulation of Tawiah *et al.* (2020) to cases without long-term survivors. They focused on patients' gap times between events and developed a joint frailty model using multivariate Gaussian random effects within a generalized linear mixed model. Unlike traditional joint shared-frailty models, this approach uses two sets of random effects to account for both intra-subject correlation in recurrent event times and individual differences in mortality hazard rates. Their formulation efficiently cancels out the unknown baseline hazard functions from the partial likelihood, making the estimation procedures relatively efficient. This method also captures the positive or negative association between recurrent and terminal events, distinguishing the origin of their dependence.

Despite the strengths presented by the aforementioned joint frailty models, the assumption of Gaussian random effects, which is traditionally the most classical approach to model dependency in repeated observations over longitudinal trajectories, poses an edge in the framework of the precision medicine. This is especially the case when the aim is to identify patients profiles that exhibit different hazards, for whom costumized targeted interventions are needed. In this perspective, in the last few decades, a more recent branch of the literature is focusing on the treatment of discretely-distributed random effects (Aitkin, 1996, 1999; Hartzel et al., 2001; Azzimonti et al., 2013; Masci et al., 2019, 2021, 2022). Most of the work in this direction has been done in the context of mixed-effects regression models for continuous responses (both univariate and multivariate) and other types of responses in the exponential family, but this approach has been recently extended in the survival analysis framework, by the introduction of discretely-distributed frailties (Caroni et al., 2010; Gasperoni et al., 2020; Cancho et al., 2020, 2021; Molina et al., 2021). The main advantage of modelling discrete random-effects is two-fold. First, it provides a new type of treatment and interpretation of the units at the highest level of the hierarchy, that are clustered into latent subpopulations. When considering events nested within patients or patients nested within healthcare providers, this approach enables the clustering of patients or healthcare providers, helping to identify characteristics such as long-term and short-term survivors or more and less successful healthcare providers. Second, the non-parametric discrete distribution of random effects allows for a more flexibly modelling of the grouping effect. It avoids imposing any parametric assumption on it and significantly simplifies the computation of the response marginal distribution, thereby avoiding potential integration issues. Despite the various proposed models and their contributions to the literature, none of them have yet explored the joint modeling of recurrent and terminal events.

Motivated by the aim of developing a tool for profiling patients and identifying latent sub-populations with varying hazard rates, in this article, we extend the joint frailty model by Ng et al. (2023) to the discrete-frailty framework. The main innovation is assuming a non-parametric bivariate discrete distribution P^* of the random terms with an unknown finite number of mass points. In addition to handling both heterogeneous susceptibility to the event risk and informative censoring, this novel approach can detect a latent structure among subjects, grouping them in sub-populations where individuals are characterized by a common frailty value. Frailty values can then be easily translated in a providers' assessment, resulting more exploitable from an interpretative point of view. Our inspiration comes from Masci et al. (2019), where a semi-parametric mixed effect model with a bivariate discretely-distributed non-parametric random term was used to perform an unsupervised classification of school sub-populations based on student performance distributions. Along with the model, we propose an estimation routine via Expectation-Maximization algorithm (Dempster et al., 1977; Bishop, 2006) and discuss the possible model design choices. Since the number of subgroups is estimated by the algorithm and is not being known a priori, the proposed model can also be interpreted as an unsupervised classification tool, encouraging exploration of similarities within grouped subjects and differences between sub-populations.

The approach developed in this paper is motivated by a study on patients with Heart Failure (HF) hospitalized in the Lombardia region of Italy. HF disease often leads to recurrent hospitalization events (Kennedy, 2001; Baraldo et al., 2013; Rogers et al., 2016; Paulon et al., 2020; Spreafico and Ieva, 2021b), which usually herald a substantial worsening of patient's survival prognosis and are terminated by death. In this context our novel approach is hence of interest for several reasons. First, it addresses informative censoring and detects sub-populations of HF patients at different risks, representing a more informative interpretation tool for medical practice. It also enables the investigation of the impact of patient-specific (time-varying) characteristics on hospital readmissions and/or mortality. Specifically, a key component of patient's

care is adherence to medication, that is the process by which patients take their medication as prescribed (Vrijens *et al.*, 2012). Proper medication adherence in HF patients can improve clinical outcomes and prevent hospitalization and reduce mortality (Ponikowski *et al.*, 2016). Specifically, we focus on HF patients receiving Angiotensin-Converting Enzyme (ACE) inhibitors, i.e., disease-modifying drugs of routine use for HF therapy (McMurray *et al.*, 2012; Yancy *et al.*, 2013; Ponikowski *et al.*, 2016), and we study the effect of time-varying adherence (Spreafico and Ieva, 2021a) to ACE therapy on both re-hospitalisations and death. Both the model and its application to HF are innovative contributions to the literature.

The remainder of the article is organized as follows. In Section 2, we introduce notation, joint frailty models and the main novelty of this work, i.e., the *Joint Model with Discretely-distributed non-parametric Frailty (JMDF)* for modelling recurrent and terminal events. Section 3 covers the estimation procedure through Expectation-Maximization algorithm and model design options. In Section 4, we apply the proposed methodology to the HF administrative database provided by *Regione Lombardia - Healthcare Division* (Regione Lombardia, 2012), comparing the results to the joint frailty model by Rondeau *et al.* (2007) and by Ng *et al.* (2023). The concludin Section 5 discusses the approach's strengths, limitations, and potential future directions. Statistical analyses were performed in the R software environment (R Core Team, 2021). In order to enhance reproducibility and validation of the research, source code is available at https://github.com/mspreafico/JMDF.

2 Methods

In Section 2.1 we introduce the notation for recurrent and terminal time-to-events considering *gap times*, i.e., times between consecutive events, and we briefly recall the joint frailty modeling proposed by Rondeau *et al.* (2007) and Ng *et al.* (2023). In section 2.2, we introduce our novel JMDF methodology.

2.1 Notation and state-of-the-art methods

2.1.1 Notation

Let us consider a cohort of N independent individuals, denoted by index i ($i = 1, ..., N$), in which each subject experiences J_i recurrent events, denoted by index j. For each subject i, let T_{ij}^R denote the gap time of recurrent event j with realization t_{ij}^R, and T_i^D the gap time from the last recurrent event to the terminal one (e.g., death) with realization t_i^D. Both event-times are subject to right censoring, whose time is denoted by C_i. For each subject i we hence observe n_i gap times, with $n_i = J_i + 1$. Define $T_{ij} = \min \left\{ T_{ij}^R, T_i^D, C_i \right\}$ $\forall j = 1, ..., n_i$ as the sequence of random variables defining observed times with realizations t_{ij}. Let δ_{ij}^R be the corresponding censoring variables for the recurrent event (1 if $T_{ij} = T_{ij}^R$, 0 otherwise), and δ_i^D be the censoring variable for the terminal event (1 if $T_{in_i} = T_i^D$, 0 otherwise). Let $\boldsymbol{X}_{ij}^R \in \mathbb{R}^{p_1}$ denotes the p_1-dimensional vector of (fixed or time-dependent) covariates associated to recurrent event j for subject i, and $\boldsymbol{X}_i^D \in \mathbb{R}^{p_2}$ be the p_2-dimensional vector of covariates associated to its terminal event. The observable data for each patient i are given by $\boldsymbol{\mathcal{O}}_i = \left\{ \left(T_{ij}, T_{ij}^R, \delta_{ij}^R, \boldsymbol{X}_{ij}^R, T_i^D, \delta_i^D, \boldsymbol{X}_i^D \right); j = 1, \ldots, n_i \right\}$. The overall set of observable gap times, censoring variables and covariates is denoted by $\boldsymbol{\Theta} = \{ \boldsymbol{\mathcal{O}}_i; i = 1, \ldots, N \}$.

2.1.2 Joint frailty model with shared log-normal random effects (Rondeau *et al.*, 2007)

To account for heterogeneity in the data due to unobserved covariates, Rondeau *et al.* (2007) proposed a joint model that included a common frailty term to the individuals for the two rates related to recurrent and terminal events. Such term is assumed to follow either a Gamma or a log-Normal distribution and acts differently on the two hazard rates by means of a parameter α. In particular, the hazard functions for the *joint log-Normal frailty model* are defined by:

$$\begin{cases} h_{ij}^R \left(t | \eta_i, \boldsymbol{X}_{ij}^R \right) = h_0^R(t) \exp \left(\boldsymbol{\beta}^T \boldsymbol{X}_{ij}^R + \eta_i \right) \\ h_i^D \left(t | \eta_i, \boldsymbol{X}_i^D \right) = h_0^D(t) \exp \left(\boldsymbol{\gamma}^T \boldsymbol{X}_i^D + \alpha \eta_i \right) \end{cases} \tag{1}$$

where $h_0^R(\cdot)$ and $h_0^D(\cdot)$ are the baseline hazard functions for recurrent and terminal events, respectively; $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the vectors of fixed-effect coefficients associated to recurrent and terminal events, respectively. The random effects are iid and distributed as $\eta_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. The parameter α determines direction of the association (if significant) between the two processes.

Parameter estimation is based on a semiparametric penalized likelihood estimation or parametric estimation on the hazard function (see Rondeau *et al.* (2007) for further details) and it is implemented in the R package `frailtypack` (Rondeau *et al.*, 2012; Król *et al.*, 2017).

2.1.3 Joint frailty model with multivariate Gaussian random effects (Ng *et al.*, 2023)

Recently, Ng *et al.* (2023) proposed a joint frailty model for recurrent and terminal events with multivariate Gaussian random effects. The authors denote by $\boldsymbol{u} = (u_1, \dots, u_N)^T$ and $\boldsymbol{v} = (v_1, \dots, v_N)^T$ the N-dimensional random vectors of u_i and v_i that represent the frailty for the i-th subject to account for intra-subject correlation of gap times of the recurrent events and individual differences in mortality hazard rate for the gap time from the last recurrent event to death, respectively. Their joint frailty model is defined as:

$$\begin{cases} h_{ij}^R \left(t_{ij}^R | u_i, \boldsymbol{X}_{ij}^R \right) = h_0^R \left(t_{ij}^R \right) \exp \left(\boldsymbol{\beta}^T \boldsymbol{X}_{ij}^R + u_i \right) \\ h_i^D \left(t_i^D | v_i, \boldsymbol{X}_i^D \right) = h_0^D \left(t_i^D \right) \exp \left(\boldsymbol{\gamma}^T \boldsymbol{X}_i^D + v_i \right) \end{cases} \tag{2}$$

where $h_0^R(\cdot)$ and $h_0^D(\cdot)$ are the baseline hazard functions for recurrent events and death, respectively; $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the vectors of fixed-effect coefficients associated to recurrent and terminal events, respectively. The frailty joint distribution is

$$[u_i, v_i] \sim \mathcal{N}_2(\boldsymbol{0}, \boldsymbol{\Sigma}) \qquad \text{with} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \theta_u^2 & \rho\theta_u\theta_v \\ \rho\theta_u\theta_v & \theta_v^2 \end{bmatrix}. \tag{3}$$

This formulation allows a well-defined, straightforward interpretation of all the parameters involved, being θ_u^2 and θ_v^2 the quantifiers of unobserved heterogeneity in the two processes, whereas ρ models the dependence between $\boldsymbol{u}$ and $\boldsymbol{v}$.

Parameter estimation is performed by maximize the sum of two components: (i) the usual Cox partial log-likelihood of failure times assuming $\boldsymbol{u}$ and $\boldsymbol{v}$ fixed, i.e., $\ell_1 (\boldsymbol{\beta}, \boldsymbol{\gamma} | \boldsymbol{u}, \boldsymbol{v})$; (ii) the logarithm of the joint probability density function of random effects $\boldsymbol{u}$ and $\boldsymbol{v}$, i.e., $\ell_2 (\boldsymbol{u}, \boldsymbol{v} | \theta_u, \theta_v, \rho)$. See Ng *et al.* (2023) for further details.

2.2 Joint Model with Discretely-distributed non-parametric Frailty (JMDF)

In devising our original methodology, we build upon the framework proposed in Ng *et al.* (2023), while introducing a bivariate non-parametric discrete distribution for the random effects. This

choice is driven by the fact that a discrete distribution of frailties not only offers an additional layer of interpretation in medical practice but also provides the opportunity to discover and analyze latent partitions within the cohort of patients under consideration.

The hazard functions in the JMDF for recurrent and terminal events are defined as:

$$\begin{cases} h_{ij}^R \left(t_{ij}^R | u_i, \boldsymbol{X}_{ij}^R \right) = h_0^R \left(t_{ij}^R \right) \exp \left(\boldsymbol{\beta}^T \boldsymbol{X}_{ij}^R + u_i \right) \\ h_i^D \left(t_i^D | v_i, \boldsymbol{X}_i^D \right) = h_0^D \left(t_i^D \right) \exp \left(\boldsymbol{\gamma}^T \boldsymbol{X}_i^D + v_i \right) \end{cases} \tag{4}$$

where $h_0^R(\cdot)$ and $h_0^D(\cdot)$ are the baseline hazard functions for recurrent and terminal events, respectively; $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the vectors of fixed-effect coefficients associated to recurrent and terminal events, respectively. Random effects u_i and v_i are assumed to be distributed according to an unknown discrete measure with a finite support in $\mathbb{R}^2$, called P^*:

$$[u, v]_i \overset{iid}{\sim} P^* \quad \forall i = 1, ..., N. \tag{5}$$

P^* can be characterized by a vector $\boldsymbol{\mathcal{P}} = (\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_K)$ of $k = 1, \ldots, K$ points $\boldsymbol{P}_k = [P_k^u, P_k^v] \in \mathbb{R}^2$ where K is unknown a priori, and a vector $\boldsymbol{w} = (w_1, w_2, \ldots, w_K)$ of relative weights. Notice that $\sum_{k=1}^K w_k = 1$ and each weight $w_k = \Pr([u, v]_i = \boldsymbol{P}_k)$ expresses the probability for each subject i (that is the second-level unit) to be assigned to a certain point k.

2.2.1 Likelihood construction.

In order to define the likelihood, we initially consider K as fixed and we introduce a set of auxiliary random variables. For each subject i, we define an auxiliary random vector $\boldsymbol{z}_i$ as follows

$$\boldsymbol{z}_i = [z_{i1} \; z_{i2} \; z_{i3} \; ... \; z_{iK}] \qquad \text{where} \qquad z_{ik} = \begin{cases} 1 & \text{if } [u, v]_i = \boldsymbol{P}_k \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Thus, each auxiliary vector is distributed according to a multivariate Bernoulli distribution of parameters $\boldsymbol{w}$.

Assuming that we have observed the realizations of such auxiliary random vectors (collected in the random matrix $\boldsymbol{\mathcal{Z}}$), we can express the full likelihood of the model as

$$\mathcal{L} \left(\boldsymbol{\Omega}; \boldsymbol{\Theta} | \boldsymbol{\mathcal{Z}} \right) = \prod_{k=1}^K \prod_{i=1}^N \left[w_k \cdot \mathcal{L}_{ik}(\boldsymbol{\Omega}; \boldsymbol{\mathcal{O}}_i | \boldsymbol{z}_i) \right]^{z_{ik}} \tag{7}$$

where $\boldsymbol{\Theta}$ is the overall set of observable data (see Section 2.1), and $\boldsymbol{\Omega} = \left[\boldsymbol{\beta}, \boldsymbol{\gamma}, H_0^R(t), H_0^D(t), \boldsymbol{w}, \boldsymbol{\mathcal{P}} \right]$ denotes the quantities to be estimated, with $H_0^R(t)$ and $H_0^D(t)$ being the cumulative baseline hazards related to the two processes. Each individual contribution to the likelihood can be written as the following product

$$\mathcal{L}_{ik} \left(\boldsymbol{\Omega}; \boldsymbol{\mathcal{O}}_i | \boldsymbol{z}_i \right) = \prod_{j=1}^{n_i} \mathcal{L}_{ijk}^R \left(\boldsymbol{\Omega}; \boldsymbol{\mathcal{O}}_{ij}^R | \boldsymbol{z}_i \right) \cdot \mathcal{L}_{ik}^D \left(\boldsymbol{\Omega}; \boldsymbol{\mathcal{O}}_i^D | \boldsymbol{z}_i \right) \tag{8}$$

with

$$\mathcal{L}_{ijk}^R \left(\boldsymbol{\Omega}; \boldsymbol{\mathcal{O}}_{ij}^R | \boldsymbol{z}_i \right) = \left[h_0^R \left(t_{ij}^R \right) \exp \left(\boldsymbol{\beta}^T \boldsymbol{X}_{ij}^R + P_k^u \right) \right]^{\delta_{ij}^R} \cdot \exp \left\{ -H_0^R \left(t_{ij}^R \right) \exp \left(\boldsymbol{\beta}^T \boldsymbol{X}_{ij}^R + P_k^u \right) \right\} \tag{9}$$

and

$$\mathcal{L}_{ik}^D \left(\boldsymbol{\Omega}; \boldsymbol{\mathcal{O}}_i^D | \boldsymbol{z}_i \right) = \left[h_0^D \left(t_i^D \right) \exp \left(\boldsymbol{\gamma}^T \boldsymbol{X}_i^D + P_l^v \right) \right]^{\delta_i^D} \cdot \exp \left\{ -H_0^D \left(t_i^D \right) \exp \left(\boldsymbol{\gamma}^T \boldsymbol{X}_i^D + P_k^v \right) \right\}, \tag{10}$$

where P_k^u and P_k^v stand for the coordinates of the chosen support point $\boldsymbol{P}_k$, respectively.

We can then express the log-likelihood $\ell\left(\boldsymbol{\Omega};\boldsymbol{\Theta}|\boldsymbol{\mathcal{Z}}\right)$ as the sum of three terms:

$$\ell\left(\boldsymbol{\Omega};\boldsymbol{\Theta}|\boldsymbol{\mathcal{Z}}\right) = \ell_w\left(\boldsymbol{\Omega}_w;\boldsymbol{\Theta}|\boldsymbol{\mathcal{Z}}\right) + \ell_R\left(\boldsymbol{\Omega}_R;\boldsymbol{\Theta}_R|\boldsymbol{\mathcal{Z}}\right) + \ell_D\left(\boldsymbol{\Omega}_D;\boldsymbol{\Theta}_D|\boldsymbol{\mathcal{Z}}\right) \tag{11}$$

where

$$\ell_w\left(\boldsymbol{\Omega}_w;\boldsymbol{\Theta}|\boldsymbol{\mathcal{Z}}\right) = \sum_{k=1}^{K}\sum_{i=1}^{N} z_{ik} \cdot \log(w_k), \tag{12}$$

$$\ell_R\left(\boldsymbol{\Omega}_R;\boldsymbol{\Theta}_R|\boldsymbol{\mathcal{Z}}\right) = \sum_{k=1}^{K}\sum_{i=1}^{N} z_{ik} \cdot \left[\sum_{j=1}^{n_i} \delta_{ij}^R \left\{\log\left(h_0^R\left(t_{ij}^R\right)\right) + \boldsymbol{\beta}^T\boldsymbol{X}_{ij}^R + P_k^u\right\}\right. \tag{13}$$
$$\left. - H_0^R\left(t_{ij}^R\right)\exp\left(\boldsymbol{\beta}^T\boldsymbol{X}_{ij}^R + P_k^u\right)\right],$$

$$\ell_D\left(\boldsymbol{\Omega}_D;\boldsymbol{\Theta}_D|\boldsymbol{\mathcal{Z}}\right) = \sum_{k=1}^{K}\sum_{i=1}^{N} z_{ik} \cdot \left[\delta_i^D \left\{\log\left(h_0^D\left(t_i^D\right)\right) + \boldsymbol{\gamma}^T\boldsymbol{X}_i^D + P_k^v\right\}\right. \tag{14}$$
$$\left. - H_0^D(t_i^D)\exp\left(\boldsymbol{\gamma}^T\boldsymbol{X}_i^D + P_k^v\right)\right].$$

Estimators for $\boldsymbol{\Omega} = [\boldsymbol{\Omega}_w, \boldsymbol{\Omega}_R, \boldsymbol{\Omega}_D]$ can be obtained by maximizing Eq. (11) using the Expectation-Maximization algorithm (Dempster *et al.*, 1977; Bishop, 2006) proposed in Section 3.1.

3 Estimation and Model design

In this section we present the Expectation-Maximization algorithm (Section 3.1) we propose to estimate parameters, discussing details on the identification of the support points (Section 3.2).

3.1 A tailored Expectation-Maximization algorithm

The log-likelihood in Eq. (11) is defined conditionally on the auxiliary random matrix $\boldsymbol{\mathcal{Z}}$. In order to maximize it, we propose a novel Expectation-Maximization algorithm (Dempster *et al.*, 1977; Bishop, 2006; Gasperoni *et al.*, 2020) to estimate $\boldsymbol{\Omega} = [\boldsymbol{\Omega}_w, \boldsymbol{\Omega}_R, \boldsymbol{\Omega}_D]$ for a given number of support mass points K.

Parameter initialization: The initial step involves parameter initialization, i.e., determining $\boldsymbol{\Omega}^{(0)}$. Firstly, the grid of support points for the discrete distributions is initialized to obtain $[\boldsymbol{\mathcal{P}}, \boldsymbol{w}]^{(0)}$ following the procedure outlined in Section 3.2. Next, two Cox-type models are fitted: one for the recurrent events with $\boldsymbol{\mathcal{P}}_u^{(0)}$ as the offset, and another for the terminal event with $\boldsymbol{\mathcal{P}}_v^{(0)}$ as the offset. The estimated parameters from these models, along with their corresponding estimated cumulative baseline hazard functions, are then used to initialize the remaining parameters, namely $\left[\boldsymbol{\beta}, \boldsymbol{\gamma}, H_0^R(t), H_0^D(t)\right]^{(0)}$.

E-step: At each iteration, the Expectation step consists of computing:

$$Q\left(\boldsymbol{\Omega}\right) = \mathbb{E}_{\boldsymbol{\mathcal{Z}}|\hat{\boldsymbol{\Omega}}}\left[\ell(\boldsymbol{\Omega};\boldsymbol{\Theta})\right] = \mathbb{E}_{\boldsymbol{\mathcal{Z}}|\hat{\boldsymbol{\Omega}}}\left[\ell_w\left(\boldsymbol{\Omega}_w;\boldsymbol{\Theta}\right)\right] + \mathbb{E}_{\boldsymbol{\mathcal{Z}}|\hat{\boldsymbol{\Omega}}}\left[\ell_R\left(\boldsymbol{\Omega}_R;\boldsymbol{\Theta}_R\right)\right] + \mathbb{E}_{\boldsymbol{\mathcal{Z}}|\hat{\boldsymbol{\Omega}}}\left[\ell_D\left(\boldsymbol{\Omega}_D;\boldsymbol{\Theta}_D\right)\right] \tag{15}$$

which is the expectation over $\mathbf{Z}$, given the current estimates of parameters $\hat{\mathbf{\Omega}}$, of the log-likelihood in Eq. (11) for the observed data $\mathbf{\Theta}$. This reduces to the computations of $\mathbb{E}\left[z_{ik}\big|\hat{\mathbf{\Omega}}, \mathbf{\mathcal{O}}_i\right]$ which we indicate as $\mathbb{Z}_{ik}$ and can be derived in closed form using Bayes' theorem:

$$\mathbb{Z}_{ik} = \frac{w_k \exp\left\{\sum_{j=1}^{n_i}\left(\delta_{ij}^R P_k^u - H_0^R(t_{ij}^R)\exp\{\boldsymbol{\beta}^T\boldsymbol{X}_{ij}^R + P_k^u\}\right) + \delta_i^D P_k^v - H_0^D(t_i^D)\exp\{\boldsymbol{\gamma}^T\boldsymbol{X}_i^D + P_k^v\}\right\}}{\sum_{m=1}^K w_m \exp\left\{\sum_{j=1}^{n_i}\left(\delta_{ij}^R P_m^u - H_0^R(t_{ij}^R)\exp\{\boldsymbol{\beta}^T\boldsymbol{X}_{ij}^R + P_m^u\}\right) + \delta_i^D P_m^v - H_0^D(t_i^D)\exp\{\boldsymbol{\gamma}^T\boldsymbol{X}_i^D + P_m^v\}\right\}}.$$
(16)

It is worth to notice that $\mathbb{Z}_{ik}$ represents the probability that subject i belongs to point k, given the current state of parameters. This allows to identify a latent partition of subjects into the K points.

M-step: The Maximization step consists of maximizing $Q\left(\mathbf{\Omega}\right)$ with respect to $\mathbf{\Omega}$, given the $\mathbb{Z}_{ik}$ obtained at the E-step. It is useful to notice that the three terms involved in the log-likelihood (11) depend on three disjoint subsets of parameters: $\mathbf{\Omega}_w = [\boldsymbol{w}]$, $\mathbf{\Omega}_R = \left[\boldsymbol{\beta}, H_0^R(t), \boldsymbol{\mathcal{P}}_u\right]$, and $\mathbf{\Omega}_D = \left[\boldsymbol{\gamma}, H_0^D(t), \boldsymbol{\mathcal{P}}_v\right]$, where $\boldsymbol{\mathcal{P}}_u$ and $\boldsymbol{\mathcal{P}}_v$ are respectively the vectors of abscissas and ordinates of the points composing the support of the discrete distribution. The maximization of $Q\left(\mathbf{\Omega}\right)$ can be carried out separately with respect to these three terms: $Q_w\left(\mathbf{\Omega}_w\right) := \mathbb{E}_{\mathbf{Z}|\hat{\mathbf{\Omega}}}\left[\ell_w\left(\mathbf{\Omega}_w; \mathbf{\Theta}\right)\right]$, $Q_R\left(\mathbf{\Omega}_R\right) := \mathbb{E}_{\mathbf{Z}|\hat{\mathbf{\Omega}}}\left[\ell_R\left(\mathbf{\Omega}_R; \mathbf{\Theta}_R\right)\right]$, and $Q_D\left(\mathbf{\Omega}_D\right) := \mathbb{E}_{\mathbf{Z}|\hat{\mathbf{\Omega}}}\left[\ell_D\left(\mathbf{\Omega}_D; \mathbf{\Theta}_D\right)\right]$.

Recalling that weights w_k must sum up to 1, the maximization of $Q_w\left(\mathbf{\Omega}_w\right)$ is a constrained optimization problem. Using Lagrangian optimization we obtain

$$\hat{w}_k = \frac{1}{N}\sum_{i=1}^N \mathbb{Z}_{ik} \quad \forall k = 1, ..., K.$$
(17)

The optimization of $Q_R\left(\mathbf{\Omega}_R\right)$ involves multiple parameters, so we adopt a multi-step approach. First, we estimate the abscissas $\boldsymbol{\mathcal{P}}_u$ of the support points, fixing $\boldsymbol{\beta}$ and $H_0^R(t)$ to their last available estimates:

$$\hat{P}_k^u = \log\left[\frac{\sum_{i=1}^N \mathbb{Z}_{ik}\sum_{j=1}^{n_i}\delta_{ij}^R}{\sum_{i=1}^N \mathbb{Z}_{ik}\sum_{j=1}^{n_i}H_0^R(t_{ij}^R)\exp\left(\boldsymbol{\beta}^T\boldsymbol{X}_{ij}^R\right)}\right] \quad \forall k = 1, ..., K.$$
(18)

By substituting $\hat{\boldsymbol{\mathcal{P}}}_u$ in Q_R and recalling that $\sum_{k=1}^K \mathbb{Z}_{ik} = 1$, we can rewrite Q_R in the following form:

$$Q_R\left(\boldsymbol{\beta}, H_0^R(t)\big|\hat{\boldsymbol{\mathcal{P}}}_u\right) = \sum_{i=1}^N\sum_{j=1}^{n_i}\left[\delta_{ij}^R \cdot \left\{\log\left(h_0^R\left(t_{ij}^R\right)\right) + \boldsymbol{\beta}^T\boldsymbol{X}_{ij}^R + \sum_{k=1}^K \mathbb{Z}_{ik}\hat{P}_k^u\right\} + \right.$$
$$\left. - H_0^R\left(t_{ij}^R\right) \cdot \left\{\sum_{k=1}^K \mathbb{Z}_{ik}\cdot\exp\left(\hat{P}_k^u\right)\right\}\cdot\exp\left(\boldsymbol{\beta}^T\boldsymbol{X}_{ij}^R\right)\right]$$
(19)

that is the usual full log-likelihood of a Cox model with known offset $\log\left[\sum_{k=1}^K \mathbb{Z}_{ik}\cdot\exp\left(\hat{P}_k^u\right)\right]$. With arguments similar to Johansen (1983) and Gasperoni *et al.* (2020), we can then compute the Breslow estimator for the cumulative baseline hazard for recurrent events as follows:

$$\hat{H}_0^R(t) = \sum_{ab:t_{ab}^R \leq t}\frac{d_{ab}^R}{\sum_{rs\in\mathcal{R}(t_{ab}^R)}\left\{\sum_{k=1}^K \mathbb{Z}_{rk}\cdot\exp\left(\hat{P}_k^u\right)\right\}\cdot\exp\left(\hat{\boldsymbol{\beta}}^T\boldsymbol{X}_{rs}^R\right)}$$
(20)

where t_{ab}^R is the time of recurrent event b for patient a, d_{ab}^R is the total number of recurrent events happening at time t_{ab}^R and $\mathcal{R}(t_{ab}^R)$ represents the recurrent-risk set at time t_{ab}^R.

Lastly, including $\hat{H}_0^R(t)$ in Eq. (19), we obtain the profile log-likelihood as a function of $\boldsymbol{\beta}$:

$$\ell_{\text{profile}}^R(\boldsymbol{\beta}) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \delta_{ij}^R \cdot \left[\boldsymbol{\beta}^T \boldsymbol{X}_{ij}^R - d_{ij}^R \cdot \log \sum_{ab \in \mathcal{R}(t_{ij}^R)} \left\{ \mathbb{Z}_{ak} \cdot \exp\left(\hat{P}_k^u \right) \right\} \cdot \exp\left(\boldsymbol{\beta}^T \boldsymbol{X}_{ab}^R \right) \right] \quad (21)$$

which is of the form of the usual partial log-likelihood in the Cox model with known offsets, so it is maximized through standard software in order to retrieve $\hat{\boldsymbol{\beta}}$.

The optimization of $Q_D\left(\boldsymbol{\Omega}_D\right)$ can be performed following the same procedure designed for $Q_R\left(\boldsymbol{\Omega}_R\right)$. Similarly, the estimates $\hat{\boldsymbol{\mathcal{P}}}_v$ and $\hat{H}_0^D(t)$ are given by:

$$\hat{P}_k^v = \log \left[\frac{\sum_{i=1}^{N} \mathbb{Z}_{ik} \cdot \delta_i^D}{\sum_{i=1}^{N} \mathbb{Z}_{ik} \cdot H_0^D(t_i^D) \exp\left(\boldsymbol{\gamma}^T \boldsymbol{X}_i^D \right)} \right] \quad \forall k = 1, ..., K; \quad (22)$$

$$\hat{H}_0^D(t) = \sum_{a:t_a^D \leq t} \frac{d_a^R}{\sum_{r \in \mathcal{R}(t_a^D)} \left\{ \sum_{k=1}^{K} \mathbb{Z}_{rk} \cdot \exp\left(\hat{P}_k^v \right) \right\} \cdot \exp\left(\hat{\boldsymbol{\gamma}}^T \boldsymbol{X}_r^D \right)} \quad (23)$$

where t_a^D is the terminal event time for patient a, d_a^D is the number of terminal events happened at t_a^D and $\mathcal{R}(t_a^D)$ is the terminal-risk set at time t_a^D. Regression parameters $\hat{\boldsymbol{\gamma}}$ can be retrieved by maximizing the following partial profile log-likelihood:

$$\ell_{\text{profile}}^D(\boldsymbol{\gamma}) = \sum_{i=1}^{N} \delta_i^D \cdot \left[\boldsymbol{\gamma}^T \boldsymbol{X}_i^D - d_i^D \cdot \log \sum_{a \in \mathcal{R}(t_i^D)} \left\{ \mathbb{Z}_{ak} \cdot \exp\left(\hat{P}_k^v \right) \right\} \cdot \exp\left(\boldsymbol{\gamma}^T \boldsymbol{X}_a^D \right) \right]. \quad (24)$$

3.1.1 Computation of standard errors.

To estimate the standard errors of the coefficients, we used the conventional `coxph` approach of Cox models from `survival` package (Therneau, 2021), where the inverse of the Hessian matrix is evaluated at the estimated coefficients. This enables us to assess the statistical significance of the parameters by calculating the corresponding Wald statistic (Collet, 2015). Similarly, we apply the same approach to obtain standard errors for the discrete random effects. $\forall k = 1, ..., K$, we compute the Hessian $\mathcal{H}$ as the second derivative of $Q_R(\boldsymbol{\Omega}_R)$ with respect to P_k^u and we evaluate it at $\hat{P}_k^u$:

$$\mathbb{E}\left[\mathcal{H}(\hat{P}_k^u) \right] = \frac{\partial^2 \boldsymbol{\Omega}_R}{\partial \mathcal{P}_u \partial \mathcal{P}_u'}(\hat{P}_k^u) = -\sum_{i=1}^{N} \mathbb{Z}_{ik} \sum_{j=1}^{n_i} H_0^R(t_{ij}^R) \exp\left(\boldsymbol{\beta}^T \boldsymbol{X}_{ij}^R \right) \exp\left(\hat{P}_k^u \right) \quad (25)$$

$$Var[\hat{P}_k^u] = \left(\mathcal{I}(\hat{P}_k^u) \right)^{-1} = \left(-\mathbb{E}\left[\mathcal{H}(\hat{P}_k^u) \right] \right)^{-1} \quad (26)$$

where $\mathcal{I}$ represents the information matrix. Equivalently, we compute the variance of $\hat{P}_k^v$.

3.2 Support points identification

Up to now, we have detailed the steps of the EM algorithm by considering the total number K of support points of the discrete distribution as known. In order to estimate the discrete distribution on $\mathbb{R}^2$, we propose a wrapper method that, given an initial grid, performs a support reduction, according to Masci et al. (2019).

Grid Initialization: The initial step involves defining a grid of M points in $\mathbb{R}^2$ that ideally covers the region believed to contain the true support of the discrete distribution (which is

unknown). This can be accomplished based on existing knowledge, such as insights from general exploratory analysis, medical expertise, or previously fitted models. Another option could be sampling from a specific distribution, such as:

(i) a *bivariate Gaussian* distribution, where weights are initialized according to the corresponding Normal density and then normalized to be unitary (Ripatti and Palmgren, 2002);

(ii) a *Uniform* distribution of points over a rectangle in $\mathbb{R}^2$, whose boundaries are defined to cover the supposed area of the true support.

To account for the algorithm's sensitivity to grid initialization, it is advisable to employ a general and non-informative initialization strategy. This helps mitigate potential misspecification issues. Additionally, in both cases, it is important to ensure that the number M of points in the initial grid is sufficient to adequately explore the designated region.

Support Reduction: Given the initial grid, the EM algorithm (Section 3.1) gradually performs the support reduction of the discrete distribution to identify $K < M$ mass points. At each E-M iteration, this reduction involves two steps.

(I) First, prior to the E-step, a specified threshold L is defined and the merging process is performed: if two points $\boldsymbol{P}_{m_1}$ and $\boldsymbol{P}_{m_2}$ are closer than L, in terms of a pre-defined distance metric, they collapse at a unique point $\boldsymbol{P}_{\bar{m}} = \left(\frac{P^u_{m_1} + P^u_{m_2}}{2}, \frac{P^v_{m_1} + P^v_{m_2}}{2} \right)$ with weight $w_{\bar{m}} = w_{m_1} + w_{m_2}$. The merging process begins with the pair having the minimum distance less than L and continues until no remaining pairs closer than the threshold remain, resulting in $\tilde{K}$ mass points. This process is sensitive to two design choices: the threshold L and the distance metric used for merging points. Specifically, the Euclidean distance offers a clear interpretation of the merging criterion based on geometric distance. However, the Manhattan distance could be beneficial when dealing with specific patterns in the hidden discrete distribution of random effects. Further insights into the definition of threshold L are discussed in Section 3.2.1. Subsequently, the E-step is executed and the $\mathbb{Z}^{(I)}_{ik}$ are computed for each individual i and remaining support point $k = 1, ..., \tilde{K}$.

(II) Prior to proceeding to the M-step, each individual i is assigned to the sub-population (i.e., mass point) k^*_i such that $k^*_i = \operatorname{argmax}_k \mathbb{Z}_{ik}$ and the support points that do not contain any individual are deleted. Let $\mathcal{K}^*$ denote the set of remaining mass-points with $|\mathcal{K}^*| = K$. When one or more mass points are deleted (i.e., $\tilde{K} \neq K$), the probabilities that individual i belongs to mass-points $k \in \mathcal{K}^*$ are re-parameterized in such a way that they sum up to 1:

$$\mathbb{Z}^{\text{new}}_{ik} = \frac{\mathbb{Z}^{\text{old}}_{ik}}{\sum_{m \in \mathcal{K}^*} \mathbb{Z}^{\text{old}}_{im}}. \tag{27}$$

Finally, the M-step is executed by considering in Eq. (17) the remaining re-parametrized probabilities $\mathbb{Z}_{ik} = \mathbb{Z}^{\text{new}}_{ik}$ with $k = 1, \ldots, K$.

The algorithm terminates when the number of masses in the discrete distribution is stable (i.e., no reduction happens in the current iteration) and the maximum difference between the components of the weights of the current and previous iteration is less than a stopping threshold *toll*, or when a predefined number of iterations *max.it* is reached. The overall procedure is summarized in panel Algorithm 1.

Algorithm 1 Estimation procedure of JMDF

Input parameters:

M: initial number of support points

init: type of initialization procedure (Gaussian or Uniform) with relative initial parameters

distance: type of distance (Euclidean, Manhattan,...)

L: value of threshold for the merging process

max.it: maximum number of iterations

toll: stopping threshold

Estimation procedure:

 1: **Grid initialization** $[\boldsymbol{\mathcal{P}}, \boldsymbol{w}]^{(0)}$ according to the *init* procedure with (at least) M support points

 2: **Parameter initialization** $\left[\boldsymbol{\beta}, \boldsymbol{\gamma}, H_0^R(t), H_0^D(t)\right]^{(0)}$ by distinct Cox-type models with offsets $\boldsymbol{\mathcal{P}}_u^{(0)}$ and $\boldsymbol{\mathcal{P}}_v^{(0)}$

 3: Set iteration $it = 0$, *converged* $= FALSE$, and $K^{(0)} = M$

 4: **while** !*converged* & $it \leq max.it$ **do**

 5:      Update iteration: $it = it + 1$

 6:      **Support Reduction I**: merge points closer than L (in terms of *distance*) by averaging the components and adding up their weights

 7:      **E-step**: compute $\mathbb{Z}_{ik}^{(it)}$ for each patient i and remaining support point $k = 1, \ldots, \tilde{K}^{(it)}$

 8:      **Support Reduction II**: extract the latent partition, delete empty support points, and re-parameterise the remaining conditional probabilities $\mathbb{Z}_{ik}^{(it),\text{new}}$ with $k = 1, \ldots, K^{(it)}$

 9:      **M-step**: update $\boldsymbol{\Omega}_w^{(it)} = [\boldsymbol{w}]^{(it)}$, $\boldsymbol{\Omega}_R^{(it)} = \left[\boldsymbol{\beta}, H_0^R(t), \boldsymbol{\mathcal{P}}_u\right]^{(it)}$, and $\boldsymbol{\Omega}_D^{(it)} = \left[\boldsymbol{\gamma}, H_0^D(t), \boldsymbol{\mathcal{P}}_v\right]^{(it)}$

10:      **if** $K^{(it)} = K^{(it-1)}$ & $\max_k \left| w_k^{(it)} - w_k^{(it-1)} \right| < $ *toll* **then**

11:          *converged* $= TRUE$

12:      **end if**

13: **end while**

14: Resulting estimates for parameters $\boldsymbol{\Omega}$ and number of mass points K are $\left[\boldsymbol{\mathcal{P}}, \boldsymbol{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, H_0^R(t), H_0^D(t)\right]^{(it)}$ and $K^{(it)}$, respectively.

3.2.1 Definition of threshold L

Defining the threshold L that determines which points will be collapsed is a crucial aspect of the estimation procedure, as it has a significant impact on the resulting discrete distribution and the identified number of masses. In general, it is advisable to set the threshold to the smallest value that captures a meaningful difference in subject classification for the specific application, taking into consideration available knowledge. From a practical standpoint, conducting a sensitivity analysis is recommended. This involves examining the behavior according to a fitting criterion (e.g., log-likelihood, AIC, classification log-likelihood) for different threshold values (and different runs for the grid initialization) to identify the most promising candidates. The choice of the fitting criterion itself will influence the sensitivity analysis. In this study, to fairly compare models with different number of masses K, we consider the classification likelihood

$$\mathcal{L}_{\text{class}}\left(\boldsymbol{\Omega}; \boldsymbol{\Theta} | \boldsymbol{\mathcal{Z}}\right) = \prod_{k=1}^{K} \prod_{i=1}^{N} \left[\mathcal{L}_{ik}(\boldsymbol{\Omega}; \boldsymbol{\mathcal{O}}_i | \boldsymbol{z}_i)\right]^{z_{ik}} \tag{28}$$

rather than the mixture one (McLachlan, 1982), and we computed the AIC and the BIC accordingly, by considering the number of parameters $g = p_1 + p_2 + K \times 2 + (K - 1)$.

4 Application

The approach developed and presented in Section 2.2 is motivated by a study of patients with Heart Failure (HF) undergoing ACE inhibitors treatment, where recurrent events of interest are hospitalizations due to HF and terminal event is death for any cause. In Sections 4.1 and 4.2, we introduce the real administrative HF database of *Regione Lombardia - Healthcare Division* (Regione Lombardia, 2012) and we present the joint model setting. Results of our method and the comparison with *joint parametric frailty models* by Rondeau *et al.* (2007) and Ng *et al.* (2023) are reported in Section 4.3 and 4.4, respectively.

4.1 Data

Administrative data were provided by *Regione Lombardia - Healthcare Division* within the research project *HFData* [HFData—RF-2009-1483329] (Regione Lombardia, 2012). The project database was built for non-paediatric residents in Lombardy (a region in northern Italy) which were hospitalized for HF from 2000 to 2012. A 5-years period from 2000 to 2005 was used in order to consider only "incident" HF patients, i.e., patients with no contacts with healthcare system in the previous five years due to HF. Each record in the dataset was related to (i) patient ordinary admission to hospital (Hospital Discharge Charts, HDC) – which contain data related to discharge date, length of stay in hospital and comorbidity conditions assessed as in Gagne *et al.* (2011), or (ii) pharmaceutical purchases (identified by their Anatomical Therapeutic Chemical (ATC) codes; see World Health Organization *et al.*, 2003) – which provide information on the number and times of drug purchases. Deaths were collected from the HDC Database (for in-hospital deaths) or Vital Statistics Regional Dataset (for out-hospital deaths). Further details regarding data extraction and selection are discussed in Mazzali *et al.* (2016).

   In this work, we focused on a representative sample of *HFData* related to 4,872 patients with their first HF discharge between January 2006 to December 2012. Overall survival was measured from the index hospitalization to the date of death or to the administrative censoring date (December 31st, 2012). To assess the effect of ACE treatment on both survival and re-hospitalizations, only subjects who experienced at least one hospitalization and one ACE purchase after the index event were selected. Demographics, comorbidities and adherence to ACE drugs were considered to adjust models. In particular, a dichotomous time-dependent variable that at each event-time indicates whether the patient was adherent to ACE therapy according to the proportion of days covered method with an 80% threshold was used (Spreafico and Ieva, 2021a).

4.1.1 Descriptive statistics.

A final cohort of $N = 2,970$ patients who underwent ACE inhibitors therapy and experienced at least one re-hospitalization was selected. At index event, median age and number of comorbidities were 74 years (IQR = [67; 80]) and 2 (IQR=[1; 3]) respectively, with a percentage of male patients equal to 58.1% (1,726 patients). Before the terminal event/censoring, median number of total re-hospitalizations was 3 (IQR = [2; 6]), with a maximum of 42 occurrences. At last hospitalization event, median age and number of comorbidities were 77 years (IQR = [70; 83]) and 3 (IQR=[2; 5]) with 1,058 patients (35.6%) adherent to ACE therapy. Median overall survival and final gap-time (i.e., time between the last recurrence and the terminal event/censoring) computed using the reverse Kaplan-Meier method by Schemper and Smith (1996) were 1,894 days (IQR=[1,365; 2,265]; about 5.1 years) and 485 (IQR=[197; 1,053]; about 1.3 years), respectively. At death/censoring event, 1,032 patients (34.7%) resulted adherent to ACE therapy and 2,139 (72.0%) patients were alive.

To proceed with the analyses, administrative data was reformatted as explained in Appendix A.

4.2 Joint frailty models for re-hospitalizations and death

In order to assess the role of patient's clinical history on both re-hospitalizations and death through the JMDF in Eq. (4), we assumed that the two instantaneous hazards for each patient depend on four explanatory variables: `sex` (*male* or *female*; time-fixed), `age` (in years; time-dependent), number of comorbidities (`ncom`; time-dependent) registered at the last known hospitalization, and the dichotomous `adherent` variable (0/1; time-dependent) indicating whether the patient is adherent or not to ACE therapy at the considered event time (Spreafico and Ieva, 2021a). The proposed JMDF for re-hospitalizations (R) and death (D) via discretely-distributed non-parametric random effects was given by

$$\begin{cases} h_{ij}^R\left(t_{ij}^R\right) = h_0^R\left(t_{ij}^R\right) \exp\left(\beta_1 \text{sex}_i + \beta_2 \text{age}_{ij} + \beta_3 \text{ncom}_{ij} + \beta_4 \text{adherent}_{ij} + u_i\right) \\ h_i^D\left(t_i^D\right) = h_0^D\left(t_i^D\right) \exp\left(\gamma_1 \text{sex}_i + \gamma_2 \text{age}_{in_i} + \gamma_3 \text{ncom}_{in_i} + \gamma_4 \text{adherent}_{in_i} + v_i\right) \end{cases} \quad (29)$$

where $h_0^R(\cdot)$ and $h_0^D(\cdot)$ are the baseline hazard functions of the re-hospitalizations and death, respectively, and $[u, v]_i$ are the random effects of the i-th patient distributed according to P^* as in Eq. (5). The vectors of parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$ and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ are respectively relative to the (time-dependent) vectors of covariates $\boldsymbol{x}_{ij}^R = \left(\text{sex}_i, \text{age}_{ij}, \text{ncom}_{ij}, \text{adherent}_{ij}\right)$ and $\boldsymbol{x}_i^D = \left(\text{sex}_i, \text{age}_{in_i}, \text{ncom}_{in_i}, \text{adherent}_{in_i}\right)$. Note that the covariate values in $\boldsymbol{x}_{ij}^R$ may vary for each event j experienced by a patient (except for sex), whereas the values in $\boldsymbol{x}_i^D$ are taken at the last gap-time n_i, representing the patient's last available measurement (see Appendix A).

The JMDF applied to *HFData* proposed in Eq. (29) was compared to the joint frailty models introduced by Rondeau *et al.* (2007) and Ng *et al.* (2023) (see Sections 2.1.2 and 2.1.3, respectively). The setting was the same as in Eq. (29), except for the random effects structure:

- Rondeau *et al.* (2007) assume shared log-Normal random effects with $u_i = \eta_i$ and $v_i = \alpha\eta_i$, where $\eta_i \sim \mathcal{N}(0, \sigma^2)$ is the patient-specific random intercept in the recurrent events process and α is the multiplicative parameter which quantifies the effect of the patient frailty on the terminal event process;

- Ng *et al.* (2023) assume bivariate Gaussian random effects as in Eq. (3).

4.3 Results of the joint models with discrete non-parametric frailty

The results of applying the JMDF in Eq. (29) to the cohort presented in Section 4.1.1 are now discussed step-by-step.

4.3.1 Grid initialization and identification of threshold L

For each initialization procedure (Gaussian and Uniform), we conducted multiple runs of the algorithm while varying the threshold values L (measured in terms of Euclidean distance) from 0.1 to 3 to determine the optimal one.

At each run with Gaussian initialization, we sampled $M = 1000$ points from a bivariate Gaussian distribution centered at the origin in $\mathbb{R}^2$ and with a diagonal variance-covariance matrix. To ensure a proper exploration of the space and prevent an overly informed distribution, we opted to double the variance estimated by the joint model with bivariate Gaussian random effect by Ng *et al.* (2023) (see Section 4.4) and enforce zero correlation in the initial grid definition. Weights were computed using the density of the bivariate Gaussian distribution under consideration, and then normalized to sum to one. For each run with Uniform initialization, we sampled a Uniform
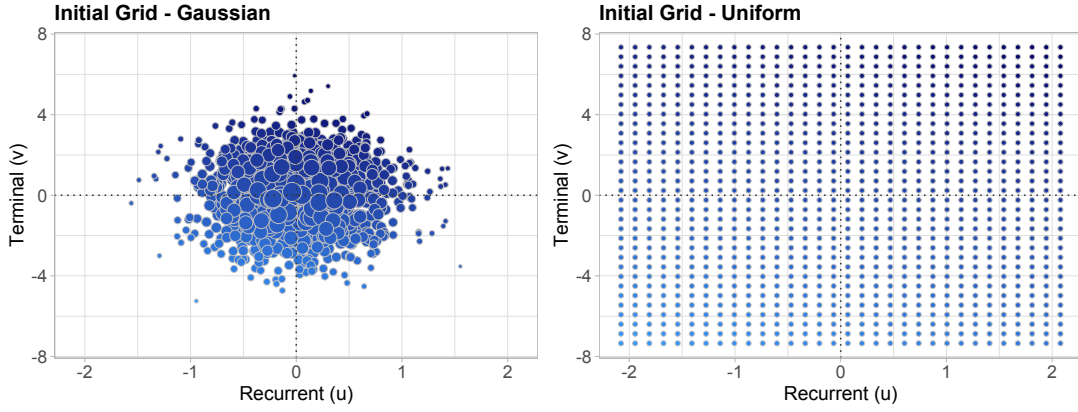
Figure 1: Gaussian (left panel) and uniform (right panel) initial grids for the random effects distribution of the discrete frailty model. Each point is colored according to a gradient scale from blue to red, which distinguishes points associated with decreased risk (blue) from one associated with increased risk (red). The size of each point reflects its weight in the discrete distribution.
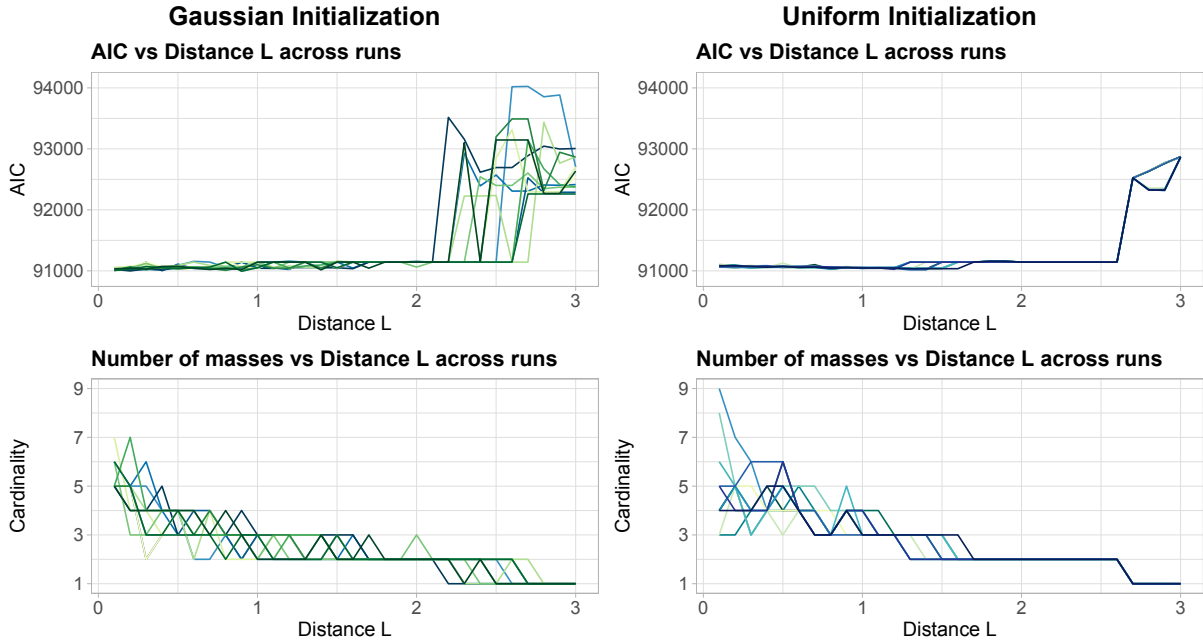

Figure 2: AIC (top panels) and number of masses (bottom panels) obtained by the joint model with discrete frailty under different distance thresholds L and across 12 runs of the algorithm. Left and right panels refer to the Gaussian and Uniform initialization cases, respectively.

distribution over a rectangle centered in the origin of $\mathbb{R}^2$ with sides' lengths set at six times the standard deviation of the corresponding variance parameters obtained by the application of the joint model by Ng *et al.* (2023). The rectangle area was filled with $M = 1024$ equally spaced points with uniform weights. Figure 1 illustrates an example of the support-point grids of the random effects for the Gaussian (left panel) and the Uniform (right-panel) initializations.

Figure 2 reports the AIC (top panels) and the number of masses K (bottom panels) of models obtained from 12 runs, each with varying values of L, using the two initialization procedures. Results on AIC showed that the Uniform case (right panel) leads to more stable and better estimates compared to the Gaussian one (right panel). This confirmed the advantage of using a more comprehensive and adaptable initialization grid. Furthermore, transitioning from the

highest number of masses (7 for the Gaussian case; 9 for the Uniform case) to 3 or even 2 masses ($L \approx 1$ or $L \approx 2$) resulted in only a slight decrease in AIC. The sensitivity analysis was repeated by computing the BIC, and the results remained consistent. In terms of mass cardinality, both the procedures highlighted the presence of stability regions for different values of L ranging from 1 to 2.5, where the fitted model suggested the presence of either 3 and 2 masses. Striking the best balance between model accuracy and complexity, these regions guided the choice of optimal threshold values. In both the Gaussian and Uniform cases, the values of L from the runs that led to the identification of $K = 2$ or $K = 3$ masses (i.e., $L \approx 1$ and $L \approx 2$, respectively) with the lowest AIC were selected as viable options.

4.3.2 Estimated mass points

Figure 3 displays the discrete random-effects estimates for the case of 2 and 3 masses, obtained with the Gaussian (left) and Uniform (right) initialization (the complete list of the estimates and their standard errors is reported in Appendix B). The disposition of points follows a linear pattern, suggesting that patients' fragility remains consistent between hospitalizations and death hazards. Notably, the range of values for the estimated random effect associated with the terminal event ($\hat{v}_i$) is broader than that of the recurrent random effect ($\hat{u}_i$). For both the initialization procedures, when 2 mass points were identified (diamonds in blue palette), the most frequent cluster (about 69% of patients) was relative to the *neutral/protected* sub-population (blue diamond $\boldsymbol{P}_1$) with a slightly negative frailty for both processes. The rest 31% can be identified as the *at-risk* sub-population (light-blue diamond $\boldsymbol{P}_2$), containing individuals at higher risks of both hospital readmission and mortality. As the value of L decreased to the point where the algorithm identified 3 mass points, we observed that the 2 clusters of patients split into 3 more differentiated clusters. This indicates that the lower value of L enables a finer separation and characterization of the patient population, revealing additional heterogeneity and subgroups within the data. In this case, results were slightly different for Gaussian and Uniform initialization but in both cases we can distinguish a *protected* sub-population ($\boldsymbol{P}_1$), a *neutral* sub-population ($\boldsymbol{P}_2$), and the sub-population *at-risk* ($\boldsymbol{P}_3$).
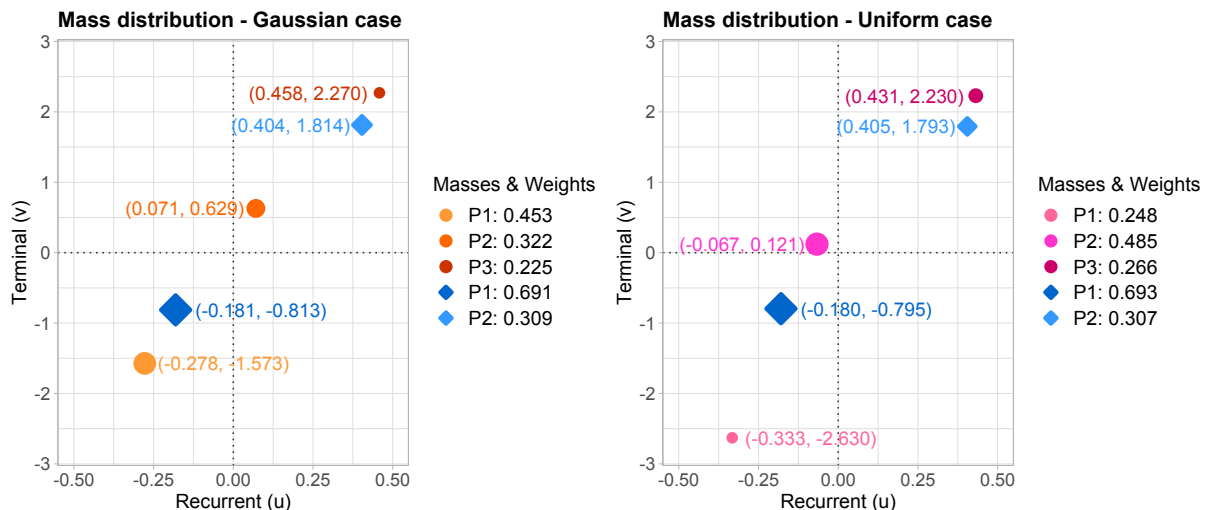

Figure 3: Estimated random-effects of the JMDF specified in Eq. (29) with Gaussian (left panel) and Uniform (right panel) initialization procedures, when $K = 3$ (circles in orange/pink palette) and $K = 2$ (diamonds in blue palette) mass points are identified. Each point is reported along with its coordinates (u, v) and the size is proportional to its mass weight w. The complete list of the estimates and their standard errors is reported in Appendix B.

This evidence highlights the strength of the proposed methodology in producing straightforward and interpretable results. A comparison of these results with the parametric continuous frailties estimated using the methods by Rondeau *et al.* (2007) and Ng *et al.* (2023) is provided in Section 4.4.

4.3.3 Estimated survival curves stratified by random effects

To better quantify and interpret the effects of belonging to different sub-populations, Figure 4 displays the estimated survival curves for a never-adherent male patient aged 74 years at baseline with two comorbidities over time. The curves are stratified by random effects for both the recurrent and terminal event processes, particularly in the more specific cases of three clusters (left panels: Gaussian; right panels: Uniform). In each panel, the black line represents the estimated survival curve for a null random effect. Regarding the re-hospitalization risk (top panels), the distinction among the three sub-populations is subtle yet evident. In both cases, the *at-risk* population exhibits a steeper curve, consistent with their shorter expected time before a new hospitalization. Aligning with the wider ranges of values for the estimated terminal-process frailties, the difference between the three sub-populations is more pronounced in terms of probability of survival (bottom panels): the *at-risk* sub-population exhibits a high mortality risk since the beginning of the follow-up, in contrast to the *neutral* sub-population, while the *protected* sub-population demonstrates good survival.
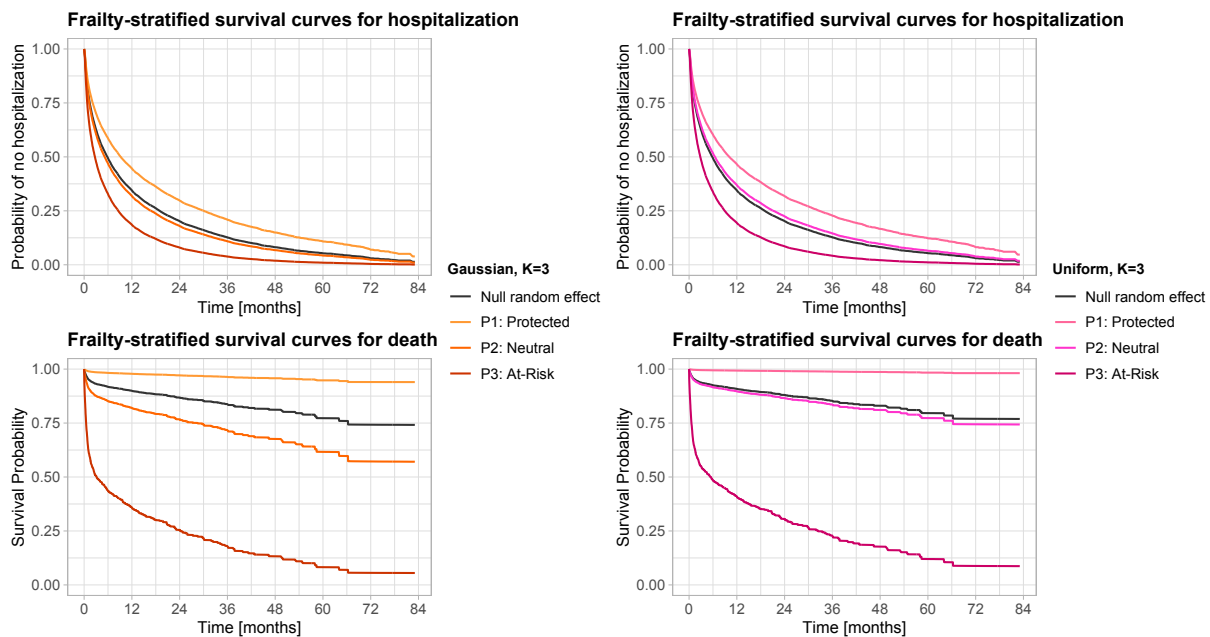

Figure 4: Estimated survival probability curves for hospitalization (top-panels) and death (bottom-panels) processes related to a never-adherent male patient aged 74 years at baseline with two comorbidities over time and stratified by random effects. The curves are associated to the discrete frailty distribution identified by using Gaussian (top-left) and Uniform (top-right) initializations when $K = 3$ masses are selected. The color of each curve is the same of the corresponding random-effect point as in Figure 3. The black lines represent the estimated survival probability curves for a null random effect. Time is expressed in days since index event.
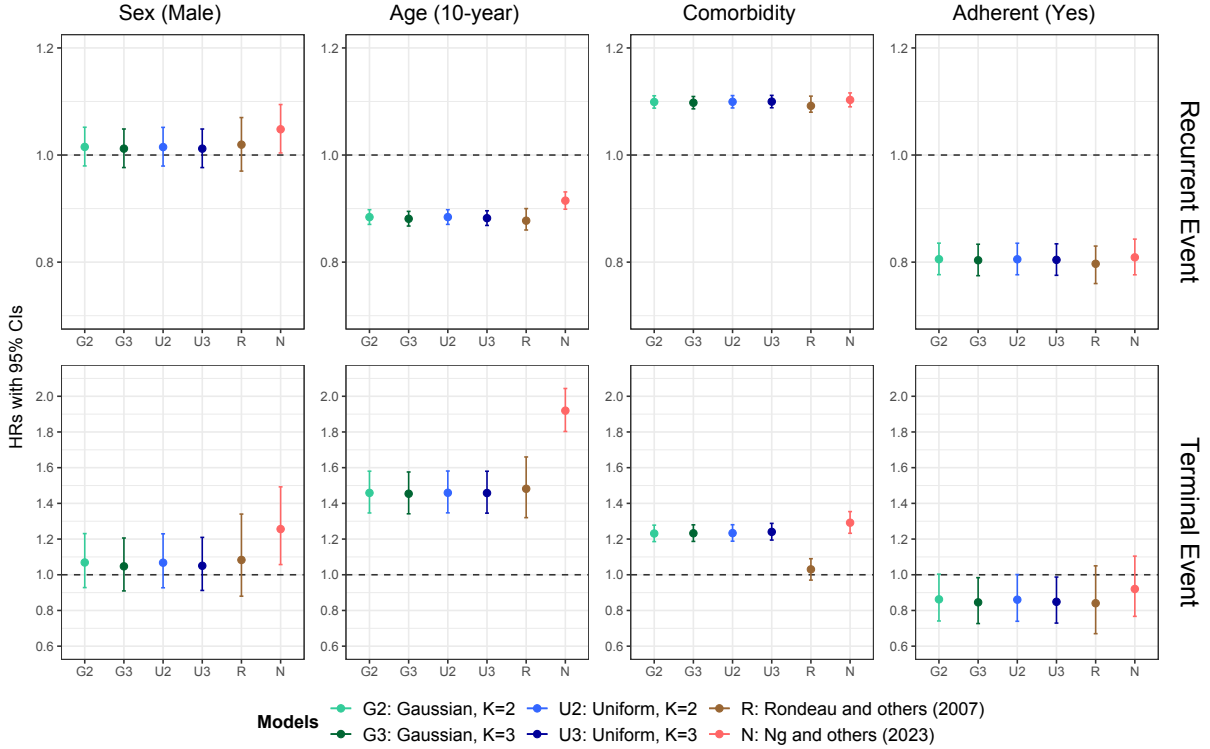
Figure 5: Comparison of estimated fixed-effect Hazard Ratios (HRs) and their 95% CIs in the trained models. Top panels refer to the recurrent hospitalization process, whereas bottom panels to terminal death process. Considered joint models are: JMDF with Gaussian initialization when $K = 2$ ($G2$: green) and $K = 3$ ($G3$: dark-green) masses are identified; JMDF with Uniform initialization when $K = 2$ ($U2$: light-blue) and $K = 3$ ($U3$: blue) masses are identified; Shared log-normal frailty model by Rondeau *et al.* (2007) (R: brown); Multivariate Gaussian frailty model by Ng *et al.* (2023) (N: salmon).

4.3.4 Effects of fixed covariates

The estimated fixed-effects (covariates `sex`, `age`, `ncom`, and `adherent`) were strongly consistent across different runs, choices of L, and initialization procedures. Figure 5 reports the Hazard Ratios (HRs) along with the 95% Confidence Intervals (CIs) for the fixed-effects estimates in both the recurrent (top panels) and terminal (bottom panels) events. Each panel displays the results for the JMDF with 2 or 3 masses, using both Gaussian and Uniform initializations, in addition to the estimates from the parametric frailty models proposed by Rondeau *et al.* (2007) and Ng *et al.* (2023). The estimated HRs showed a high level of concordance across the models, especially for the recurrent events process. This suggests that, despite varying assumptions about the random effects, the fixed-effects estimates remain consistent across the models. Another positive observation pertains to the standard errors, which are smaller in the JMDF compared to the others.

The `sex` variable (first columns) did not emerge as a significant predictor for either the hazard of recurrent or terminal events. Patient `age` was found to be statistically significant for both processes. Its effect on the hospitalization hazard is a 12% reduction in the hazard of hospitalization per 10-year increase ($HR = 0.88$). Conversely, a 10-year increase results in a 45% increase in the death hazard (HR=1.45). From a clinical perspective, this phenomenon can be explained by the fact that as patients get older, the risk of experiencing a new hospitalization is partially replaced by the risk of mortality. The number of comorbidities `ncom` resulted a statistically significant risk factor for both processes, leading to a 10% increase in the risk of hospitalization

and a substantial 23% increase in the risk of death per registered comorbidity ($HR = 1.10$ and $HR = 1.23$, respectively). This confirmed the well-documented role of comorbidities in increasing mortality and hospitalizations among HF patients (Van der Wal et al., 2017; Widmer, 2011). Finally, being `adherent` to the ACE treatment was found to be statistically significant at any level for the recurrent event process, yielding a 20% decrease in the hazard of a new hospitalization ($HR = 0.80$). For the death process, it was significant only in the 3-mass cases, where it led to a 15% decrease in the death hazard ($HR = 0.85$). From a clinical perspective, these results endorse the efficacy of ACE inhibitors treatment for HF. It demonstrates a significant reduction in the hospitalization rate, consequently lowering the occurrence of critical HF events in adherent patients throughout their clinical journey, while also enhancing their survival probability.

4.4 Comparison of random effects with joint parametric frailty models

To compare our proposed method with discrete non-parametric random effects to its counterparts, we applied the joint parametric frailty models presented by Rondeau et al. (2007) and Ng et al. (2023) to the *HFData*.

Figure 6 displays pointwise estimates for models by Rondeau et al. (2007) (left panel) and Ng et al. (2023) (right panel). In the right panel, each point corresponds to a different subject i with recurrent $\hat{u}_i$ as abscissa and terminal $\hat{v}_i$ as ordinata, and it is colored according to the assigned discrete point-mass group for Uniform initialization when $K = 3$ masses are identified as in Figure 3. In the left panel, we report the estimated $\exp(\eta_i)$ stratified according to the assigned 3 discrete mass points. Results are in line with the role of the *protected/neutral/at-risk* sub-populations.

Table 1 reports the estimated parameters for the random-effects in Figure 6. In both models, the frailty associated with the recurrent events process exhibits lower variability compared to the frailty associated with the death process. This observation aligns with the estimated distribution of our mass points (see Figure 3) and is clinically reasonable, as subjective factors influencing mortality outcomes may exhibit greater diversity and significance compared to those affecting hospitalizations. The notably high positive value estimated for the multiplicative parameter in Rondeau et al. (2007) ($\hat{\alpha} = 5.552$) and the strong positive correlation ($\hat{\rho} = 0.883$) between the frailties estimated by Ng et al. (2023) indicate a significant positive correlation between the frailties associated with the two processes, affirming the positions of the mass points in our discrete distribution. This finding emphasizes the importance of modeling the dependence between the frailties of the recurrent events process and the death process. It highlights that individuals more susceptible to re-hospitalizations are also more likely to experience higher mortality risks. By incorporating and accounting for this dependence in the model, we can better capture the interplay and shared underlying factors between these two processes.

Joint model	Random-effect Parameters	Estimate	StdDev	pvalue
Rondeau et al. (2007)	σ^2	0.127	0.009	<2e-16
	α	5.552	0.318	<2e-16
	θ_u^2	0.094	0.004	<2e-16
Ng et al. (2023)	θ_v^2	1.438	0.059	<2e-16
	ρ	0.883	0.006	<2e-16

Table 1: Estimated random-effects coefficients of the joint parametric frailty models by Rondeau et al. (2007) and Ng et al. (2023).
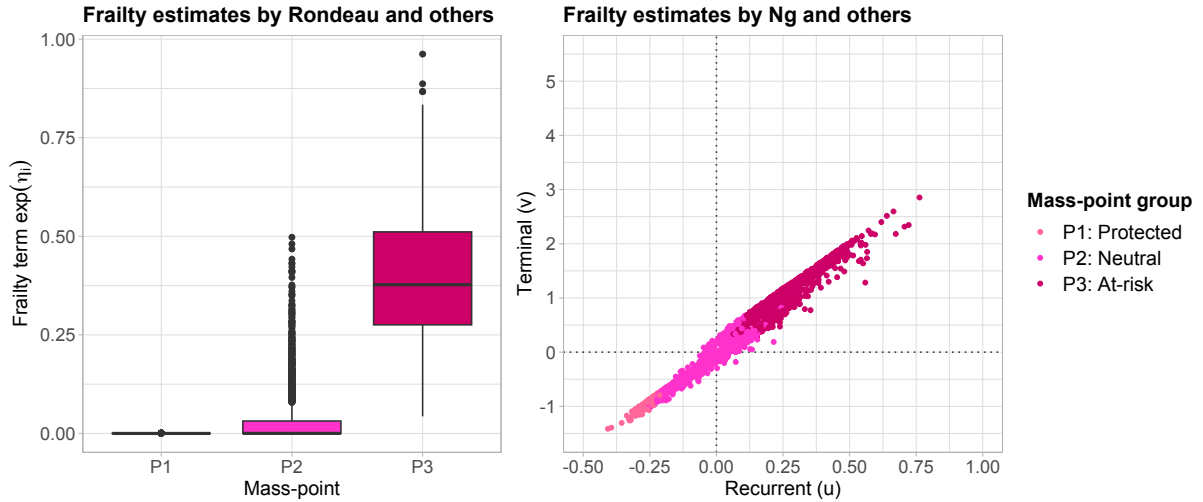
Figure 6: Random effects pointwise estimates for models by Rondeau *et al.* (2007) (left panel) and Ng *et al.* (2023) (right panel). In the right panel, points are visualized in $\mathbb{R}^2$ considering hospitalization frailties as abscissa and death frailties as ordinata. Each point correspond to a different subject and it is colored according to the assigned discrete point-mass group for Uniform initialization when $K = 3$ masses are identified as in Figure 3.In the left panel, the distribution of $\exp(\eta_i)$ is stratified according to the assigned 3 discrete mass points.

5   Discussion

This paper contributes to the literature on joint models for recurrent and terminal events by introducing an innovative joint frailty model, called JMDF, in which the frailties related to the two processes assume a discrete distribution with an a priori unknown number of support points. The JMDF approach allows for the clustering of the highest-level units, i.e., patients in our case, based on their frailty levels. After adjusting for the observable factors, the bivariate discrete frailty takes into account the heterogeneity at the patient level, associated with unobserved covariates, and captures the correlation between the two processes. Specifically, the assumption of a discrete distribution translates this heterogeneity into the identification of clusters of patients, enabling us to distinguish different patients profiles based on their associated frailty values.

The advantage of this modelling approach is twofold. First, using a discrete distribution with an unknown number of support points for frailties can indeed increase the model's flexibility compared to classic parametric distributions. Parametric distributions impose specific assumption on the shape of the frailty that might not always hold in real-world data. By using a discrete distribution, the frailties can encompass a broader range of values and patterns, potentially reducing the risk of bias and leading to a better fit to the data. Second, discrete frailties capture different patterns of vulnerability or risk among patients, with each support point corresponding to a distinct subpopulation of patients sharing a similar frailty profile. These subpopulations represent individuals who share unobserved common characteristics, behaviors, or risk factors affecting their outcomes. This lends interpretability to the subpopulations and provides insights for tailoring interventions to address the specific needs of each group.

When applied to *HFData* for modelling the two correlated processes related to hospitalizations and death, the JMDF suggested the presence of three different patients subpopulations, namely the *protected*, *neutral*, and *at-risk* populations. Compared to the methods proposed in Ng *et al.* (2023) and Rondeau *et al.* (2007), our fixed-effects estimates aligned with those estimated by the counterparts and, moreover, exhibited smaller standard errors. Regarding random effects, the subpopulations identified by JMDF were consistent with the distribution of contin-

uous frailties estimated by Ng *et al.* (2023) and Rondeau *et al.* (2007). These findings support the reliability of the proposed method and its results.

Alongside these advantages, our approach has also some limitations and possibilities for future developments. Firstly, the estimation procedure is highly sensitive to the choice of the parameter L, which tunes the spacing between mass points in the frailty discrete distribution. When users have *a priori* knowledge about the magnitude of differences they aim to observe across patient clusters, the parameter L represents a positive key point. Otherwise, a sensitivity analysis is necessary to address this issue and identify potential values for L. By exploiting the clusters identified using different L values, we can identify stability regions and compare their goodness-of-fit indices, achieving a balance between model complexity and the ability to capture meaningful patterns. This procedure also allows for result evaluation at different granularity levels, revealing varying degrees of heterogeneity among the subpopulations. Nonetheless, tuning this parameter might be time-consuming and challenging. Further work will be devoted on developing a significance-based method in which the identified clusters differ in terms of statistical significance (Ragni *et al.*, 2023). Second, the type of discrete distribution we assume assigns each patient to a cluster that describes the dynamic related to both the death and hospitalization processes. The two processes are assumed to arise from a distribution with the same number of mass points. However, this assumption may not always reflect the underlying reality as there might be patients sharing the same type of hospitalization process but not the same type of death process and vice-versa. Patient-level heterogeneity can vary between these processes, and the optimal number of clusters to capture the dynamics within the data could differ as well. Therefore, relaxing this assumption and using a more flexible discrete distribution present a promising avenue for future research in this field. By allowing for varying numbers of clusters or mass points in the frailty distribution for each process, researchers can potentially provide a more accurate and nuanced representation of the underlying phenomena.

Overall, this work enriches the literature on joint frailty models for recurrent and terminal events by embracing discretely-distributed non-parametric frailties. This new methodology empowers the identification of subgroups of patients united by shared frailty attributes. In collaboration with healthcare professionals, this additional information has the potential to better profile patients and, in turn, improve the refinement of their therapeutic pathways.

References

Aalen, O. O., Borgan, O., and Gjessing, H. K. (2008). *Survival and event history analysis.* Statistics for Biology and Health. Springer, 2008 edition.

Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and computing*, 6(3):251–262.

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, 55(1):117–128.

Amorim, L. D. A. F. and Cai, J. (2015). Modelling recurrent events: a tutorial for analysis in epidemiology. *International journal of epidemiology*, 44(1):324–333.

Azzimonti, L., Ieva, F., and Paganoni, A. M. (2013). Nonlinear nonparametric mixed-effects models for unsupervised classification. *Computational Statistics*, 28(4):1549–1570.

Baraldo, S., Ieva, F., Paganoni, A. M., and Vitelli, V. (2013). Outcome prediction for heart failure tele-monitoring via generalized linear models with functional covariates. *Scandinavian Journal of Statistics*, 40(3):403–416.

Bishop, P. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Pringer-Verlag.

Cancho, V. G., Barriga, G., Leão, J., and Saulo, H. (2021). Survival model induced by discrete frailty for modeling of lifetime data with long-term survivors and change-point. *Communication Statistics - Theory Methods*, 50(5):1161–1172.

Cancho, V. G., Suzuki, A. K., Barriga, G. D. C., and do Espirito Santo, A. P. J. (2020). A multivariate survival model induced by discrete frailty. *Communications in Statistics - Simulation and Computation*, pages 1–19.

Caroni, C., Crowder, M., and Kimber, A. (2010). Proportional hazards models with discrete frailty. *Lifetime Data Analysis*, 16(3):374–384.

Chauvet, J. and Rondeau, V. (2023). A flexible class of generalized joint frailty models for the analysis of survival endpoints. *Statistics in Medicine*, 42(8):1233–1262.

Choi, Y.-H., Jacqmin-Gadda, H., Król, A., Parfrey, P., Briollais, L., and Rondeau, V. (2020). Joint nested frailty models for clustered recurrent and terminal events: An application to colonoscopy screening visits and colorectal cancer risks in Lynch Syndrome families. *Statistical Methods in Medical Research*, 29(5):1466–1479.

Collet, D. (2015). *Modelling survival data in medical research*. Chapman & Hall.

Cook, R. J. and Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*. New York: Springer.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.

Gagne, J. J., Glynn, R. J., Avorn, J., Levin, R., and Schneeweiss, S. (2011). A combined comorbidity score predicted mortality in elderly patients better than existing scores. *Journal of Clinical Epidemiology*, 64(7):749–759.

Gasperoni, F., Ieva, F., Paganoni, A., Jackson, C., and Sharples, L. (2020). Non-parametric frailty cox models for hierarchical time-to-event data. *Biostatistics*, 21(3):531—-544.

Hartzel, J., Agresti, A., and Caffo, B. (2001). Multinomial logit random effects models. *Statistical Modelling*, 1(2):81–102.

Hougaard, P. (1995). Frailty models for survival data. *Lifetime data analysis*, 1(3):255–273.

Hougaard, P. (2012). *Analysis of multivariate survival data*. Statistics for Biology and Health. Springer, New York, NY.

Huang, X. and Liu, L. (2007). A joint frailty model for survival and gap times between recurrent events. *Biometrics*, 63(2):389–397.

Huang, X. and Wolfe, R. (2002). A frailty model for informative censoring. *Biometrics*, 58:510–520.

Johansen, S. (1983). An extension of cox's regression model. *International Statistical Review / Revue Internationale de Statistique*, 51(2):165–174.

Kennedy, B. (2001). Repeated hospitalizations and self-rated health among the elderly: A multivariate failure time analysis. *American Journal of Epidemiology*, 153:232–241.

Kleinbaum, D. G. and Klein, M. (1996). *Survival Analysis: A Self-Learning Text*. Springer.

Król, A., Mauguen, A., Mazroui, Y., Laurent, A., Michiels, S., and Rondeau, V. (2017). Tutorial in joint modeling and prediction: A statistical software for correlated longitudinal outcomes, recurrent events and a terminal event. *Journal of Statistical Software*, 81(3):1–52.

Lancaster, T. and Intrator, O. (1998). Panel data with survival: Hospitalization of hiv-positive patients. *Journal of the American Statistical Association*, 93:46–53.

Liu, L., Huang, X., Yaroshinsky, A., and Cormier, J. N. (2016). Joint frailty models for zero-inflated recurrent events in the presence of a terminal event. *Biometrics*, 72(1):204–214.

Liu, L., Wolfe, R. A., and Huang, X. (2004). Shared Frailty Models for Recurrent Events and a Terminal Event. *Biometrics*, 60(3):747–756.

Masci, C., Ieva, F., Agasisti, T., and Paganoni, A. M. (2021). Evaluating class and school effects on the joint student achievements in different subjects: A bivariate semiparametric model with random coefficients. *Computational Statistics*, 36(4):2337–2377.

Masci, C., Ieva, F., and Paganoni, A. M. (2022). Semiparametric multinomial mixed-effects models: A university students profiling tool. *The Annals of Applied Statistics*, 16(3):1608–1632.

Masci, C., Paganoni, A. M., and Ieva, F. (2019). Semiparametric mixed effects models for unsupervised classification of italian schools. *Journal of the Royal Statistical Society Series A*, 182:1313–1342.

Mazzali, C., Paganoni, A. M., Ieva, F., Masella, C., Maistrello, M., Agostoni, O., Scalvini, S., Frigerio, M., and On behalf of the HF Data Project (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).

McLachlan, G. J. (1982). 9 the classification and mixture maximum likelihood approaches to cluster analysis. *Handbook of statistics*, 2:199–208.

McMurray, J. J., Adamopoulos, S., Anker, S. D., Auricchio, A., Böhm, M., Dickstein, K., Falk, V., Filippatos, G., Fonseca, C., Gomez-Sanchez, M. A., Jaarsma, T., Køber, L., Lip, G., Maggioni, A., Parkhomenko, A., Pieske, B., Popescu, B., Rønnevik, P., Rutten, F., Schwitter, J., Seferovic, P., Stepinska, J., Trindade, P., Voors, A., Zannad, F., and Zeiher, A. (2012). ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treat-ment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *European Heart Journal*, 14:803–869.

Molina, K. C., Calsavara, V. F., Tomazella, V. D., and Milani, E. A. (2021). Survival models induced by zero-modified power series discrete frailty: Application with a melanoma data set. *Statistical Methods in Medical Research*, 30(8):1874–1889.

Ng, S. K., Tawiah, R., McLachlan, G. J., and Gopalan, V. (2023). Joint frailty modeling of time-to-event data to elicit the evolution pathway of events: a generalized linear mixed model approach. *Biostatistics*, 24(1):108–123.

Ozga, A.-K., Kieser, M., and Rauch, G. (2018). A systematic comparison of recurrent event models for application to composite endpoints. *BMC medical research methodology*, 18(1):1–12.

Paulon, G., De Iorio, M., Guglielmi, A., and Ieva, F. (2020). Joint modeling of recurrent events and survival: a bayesian non-parametric approach. *Biostatistics*, 21:1–14.

Ponikowski, P., Voors, A. A., Anker, S. D., Bueno, H., Cleland, J., Coats, A., Falk, V., González-Juanatey, J, R., Harjola, V., Jankowska, E., Jessup, M., Linde, C., Nihoyannopoulos, P., Parissis, J., Pieske, B., Riley, J., Rosano, G., Ruilope, L., Ruschitzka, F., Rutten, F., van der Meer, P., and Group, E. S. D. (2016). 2016 esc guidelines for the diagnosis and treatment of acute and chronic heart failure: The

task force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *European Heart Journal*, 37(27):2129–2200.

R Core Team (2021). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ragni, A., Masci, C., Ieva, F., and Paganoni, A. M. (2023). Clustering hierarchies via a semi-parametric generalized linear mixed model: a statistical significance-based approach. *arXiv preprint arXiv:2302.12103*.

Regione Lombardia (2012). *HFData project: Utilization of Regional Health Source Databases for Evaluating Epidemiology, short- and medium-term outcome and process indicators in patients hospitalized for heart failure.* Progetto di Ricerca Finalizzata di Regione Lombardia - HFData-RF-2009-1483329.

Ripatti, S. and Palmgren, J. (2002). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56:1016–1022.

Rogers, J. K., Yaroshinsky, A., Pocock, S. J., Stokar, D., and Pogoda, J. (2016). Analysis of recurrent events with an associated informative dropout time: Application of the joint frailty model. *Statistics in Medicine*, 35(13):2195–2205.

Rondeau, V., Commenges, D., and Joly, P. (2003). Maximum Penalized Likelihood Estimation in a Gamma-Frailty Model. *Lifetime Data Analysis*, 9:139–153.

Rondeau, V., Filleul, L., and Joly, P. (2006). Nested frailty models using maximum penalized likelihood estimation. *Statistics in Medicine*, 25(23):4036–4052.

Rondeau, V., Marzroui, 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. *Journal of Statistical Software*, 47(4):1–28.

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. *International Journal of Epidemiology*, 8:708–721.

Schemper, M. and Smith, T. L. (1996). A note on quantifying follow-up in studies of failure time. *Controlled Clinical Trials*, 17(4):343–346.

Spreafico, M. and Ieva, F. (2021a). Dynamic monitoring of the effects of adherence to medication on survival in heart failure patients: A joint modeling approach exploiting time-varying covariates. *Biometrical Journal*, 63(2):305–322.

Spreafico, M. and Ieva, F. (2021b). Functional modeling of recurrent events on time-to-event processes. *Biometrical Journal*, 63(5):948–967.

Talebi-Ghane, E., Baghestani, A., Zayeri, F., Rondeau, V., and Akhavan, A. (2021). Joint frailty model for recurrent events and death in presence of cure fraction: Application to breast cancer data. *Biometrical Journal*, 63(4):725–744.

Tawiah, R., McLachlan, G. J., and Ng, S. K. (2020). A bivariate joint frailty model with mixture framework for survival analysis of recurrent events with dependent censoring and cure fraction. *Biometrics*, 76(3):753–766.

Therneau, T. M. and Grambsch, P. (2000). *Modeling Survival Data: Extending the Cox Model.* Springer, New York.

Therneau, T. (2021). *A Package for Survival Analysis in R.* R package version 3.2-13. https://CRAN.R-project.org/package=survival>.

Van der Wal, H., Van Deursen, V., Van der Meer, P., and Voors, A. (2017). Comorbidities in heart failure. *Handbook of Experimental Pharmacology*, 243:35–66.

Vrijens, B., De Geest, S., Hughes, D. A., Przemyslaw, K., Demonceau, J., Ruppar, T., Dobbels, F., Fargher, E., Morrison, V., Lewek, P., Matyjaszczyk, M., Mshelia, C., Clyne, W., Aronson, J. K., Urquhart, J., and ABC Project Team (2012). A new taxonomy for describing and defining adherence to medications. *British Journal of Clinical Pharmacology*, 73(5):691–705.

Widmer, F. (2011). Herzinsuffizienz und komorbiditäten – Comorbidity in heart failure. *Therapeutische Umschau*, 68(2):103–106.

World Health Organization, WHO International Working Group for Drug Statistics Methodology, WHO Collaborating Centre for Drug Statistics Methodology, and WHO Collaborating Centre for Drug Utilization Research and Clinical Pharmacological Service (2003). *Introduction to Drug Utilization Research*. WHO Library Cataloguing-in-Publication Data.

Yancy, C. W., Jessup, M., Bozkurt, B., Butler, J., Casey, Jr, D. E., Drazner, M. H., Fonarow, G. C., Geraci, S. A., Horwich, T., Januzzi, J. L., Johnson, M. R., Kasper, E. K., Levy, W. C., Masoudi, F. A., McBride, P. E., McMurray, J. J. V., Mitchell, J. E., Peterson, P. N., Riegel, B., Sam, F., Stevenson, L. W., Tang, W. H. W., Tsai, E. J., and Wilkoff, B. L. (2013). 2013 ACCF/AHA guideline for the management of heart failure: executive summary: a report of the american college of cardiology Foundation/American heart association task force on practice guidelines. *Circulation*, 128(16):1810–1852.

Zeng, D. and Lin, D. Y. (2009). Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. *Biometrics*, 65(3):746–752.

Appendix

A   Data format

After selecting the cohort of patients for analysis and identifying the relevant events in each patient's clinical history (see Section 4.1), we proceeded to reformat the administrative data to adhere to the required format for the `coxph` function in the `survival` R package (Therneau, 2021). Table 2 shows an example of reformatted dataset related to two hypothetical patients $i \in \{A, B\}$.

The dataset contains 11 rows, which corresponds to the sum of the number of gap times T_{ij} for both patient A and patient B, i.e., $n_A = 5$ and $n_B = 6$ respectively. Patient A experienced four re-hospitalization events ($\delta^R_{Aj} = 1$ and $\delta^D_{Aj} = 0$ for $j = 1, \ldots, 4$), and was censored at the last follow-up ($\delta^D_{A5} = 0$ and $\delta^R_{A5} = 0$). Patient B experienced five re-hospitalization events ($\delta^R_{Bj} = 1$ and $\delta^D_{Bj} = 0$ for $j = 1, \ldots, 5$), and died at the last follow-up ($\delta^D_{B6} = 1$ and $\delta^R_{A6} = 0$).

The dataset contains four explanatory variables: patient's $\texttt{sex}_i$ (*male* or *female*; time-fixed), time-dependent $\texttt{age}_{ij}$ (in years) and number of comorbidities ($\texttt{ncom}_{ij}$) registered at the last known hospitalization, and time-dependent dichotomous variable $\texttt{adherent}_{ij}$ indicating whether the patient was adherent to ACE therapy according to the proportion of days covered method with an 80% threshold (Spreafico and Ieva, 2021a).

The (time-dependent) vectors of covariates $\boldsymbol{x}^R_{ij} = \left(\texttt{sex}_i, \texttt{age}_{ij}, \texttt{ncom}_{ij}, \texttt{adherent}_{ij}\right)$ and $\boldsymbol{x}^D_i = \left(\texttt{sex}_i, \texttt{age}_{in_i}, \texttt{ncom}_{in_i}, \texttt{adherent}_{in_i}\right)$. Note that the covariate values in $\boldsymbol{x}^R_{ij}$ may vary for each event j experienced by a patient (except for sex), whereas the values in $\boldsymbol{x}^D_i$ are taken at the last gap-time n_i, representing the patient's last available measurement. As an example, considering the second and the last gap times for each patient in Table 2, we have the following values:

$$\boldsymbol{x}^R_{A2} = (\textit{female}, 67, 5, 1) \qquad \text{and} \qquad \boldsymbol{x}^R_{B2} = (\textit{male}, 77, 2, 0);$$

$$\boldsymbol{x}^D_A = \boldsymbol{x}^R_{A6} = (\textit{female}, 71, 5, 0) \qquad \text{and} \qquad \boldsymbol{x}^D_B = \boldsymbol{x}^R_{B6} = (\textit{male}, 79, 4, 1).$$

i	j	δ^R_{ij}	δ^D_{ij}	T_{ij}	$\texttt{sex}_i$	$\texttt{age}_{ij}$	$\texttt{ncom}_{ij}$	$\texttt{adherent}_{ij}$
A	1	1	0	49	*female*	65	5	1
A	2	1	0	901	*female*	67	5	1
A	3	1	0	391	*female*	69	5	1
A	4	1	0	10	*female*	69	5	1
A	5	0	0	801	*female*	71	5	0
B	1	1	0	82	*male*	77	2	0
B	2	1	0	11	*male*	77	2	0
B	3	1	0	186	*male*	77	2	1
B	4	1	0	29	*male*	77	2	1
B	5	1	0	118	*male*	78	4	1
B	6	0	1	183	*male*	79	4	1

Table 2: Example of reformatted dataset.

B Discrete frailty estimates

		Gaussian initialization			Uniform initialization		
		$\widehat{P}_k^u$ (sd)	$\widehat{P}_k^v$ (sd)	$\widehat{w}_k$	$\widehat{P}_k^u$ (sd)	$\widehat{P}_k^v$ (sd)	$\widehat{w}_k$
$K = 3$	$\boldsymbol{P}_1$	-0.278 (0.0002)	-1.573 (0.0199)	0.453	-0.333 (0.0004)	-2.630 (0.1131)	0.248
	$\boldsymbol{P}_2$	0.071 (0.0002)	0.629 (0.0036)	0.322	-0.067 (0.0002)	0.121 (0.0040)	0.485
	$\boldsymbol{P}_3$	0.458 (0.0002)	2.270 (0.0019)	0.225	0.431 (0.0002)	2.230 (0.0017)	0.266
$K = 2$	$\boldsymbol{P}_1$	-0.181 (0.0001)	-0.813 (0.0051)	0.691	-0.180 (0.0001)	-0.795 (0.0049)	0.693
	$\boldsymbol{P}_2$	0.404 (0.0002)	1.814 (0.0016)	0.309	0.405 (0.0002)	1.793 (0.0016)	0.307

Table 3: Estimated random-effects of the JMDF specified in Eq. (29) with Gaussian and Uniform initialization procedures, when $K = 3$ and $K = 2$ mass points are identified. Each point k is reported in terms of its coordinates $\widehat{\boldsymbol{P}}_k = \left(\widehat{P}_k^u, \widehat{P}_k^v \right)$ and its weight $\widehat{w}_k$. Standard errors of the estimates are reported in brackets.

MOX Technical Reports, last issues

Dipartimento di Matematica

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

87/2023 Masci, C.; Spreafico, M.; Ieva, F.
Joint modelling of recurrent and terminal events with discretely-distributed non-parametric frailty: application on re-hospitalizations and death in heart failure patients

85/2023 Arnone, E.; De Falco, C.; Formaggia, L.; Meretti, G.; Sangalli, L.M.
Computationally efficient techniques for Spatial Regression with Differential Regularization

86/2023 Ferraccioli, F.; Sangalli, L.M.; Finos, L.
Nonparametric tests for semiparametric regression models

83/2023 Cavinato, L.; Massi, M.C.; Sollini, M.; Kirienko , M.; Ieva, F.
Dual adversarial deconfounding autoencoder for joint batch-effects removal from multi-center and multi-scanner radiomics data

82/2023 Pozzi, G.; Ciarletta, P.
Geometric control by active mechanics of epithelial gap closure

81/2023 Buchwald, S.; Ciaramella, G; Salomon, J.; Sugny, D.
A SPIRED code for the reconstruction of spin distribution

79/2023 Agosti, A.; Bardin, R.; Ciarletta, P.; Grasselli, M.
A diffuse interface model of tumour evolution under a finite elastic confinement

78/2023 Antonietti, P.F.; Bonizzoni, F.; Corti, M.; Dall'Olio, A.
Discontinuous Galerkin for the heterodimer model of prion dynamics in Parkinson's disease

77/2023 Fumagalli, I.; Corti, M.; Parolini, N.; Antonietti, P. F.
Polytopal discontinuous Galerkin discretization of brain multiphysics flow dynamics

76/2023 Ieva, F.; Galliani, G.; Secchi, P.
The impact of public transport on the diffusion of COVID-19 pandemie in Lombardy during 2020