MOX-Report No. 04/2026

Penalized Likelihood Optimization for Adaptive Neighborhood Clustering in Time-to-Event Data with Group-Level Heterogeneity

Ragni, A.; Cavinato, L.; Ieva, F.

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

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

Penalized Likelihood Optimization for Adaptive Neighborhood Clustering in Time-to-Event Data with Group-Level Heterogeneity

Alessandra Ragni[1], Lara Cavinato[1], Francesca Ieva[1,2]

[1]MOX, Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133, Milano, Italy

[2]Health Data Science Centre, Human Technopole, Viale Rita Levi–Montalcini 1, Milan 20157, Italy

Abstract

The identification of patient subgroups with comparable event-risk dynamics plays a key role in supporting informed decision-making in clinical research. In such settings, it is important to account for the inherent dependence that arises when individuals are nested within higher-level units, such as hospitals. Existing survival models account for group-level heterogeneity through frailty terms but do not uncover latent patient subgroups, while most clustering methods ignore hierarchical structure and are not estimated jointly with survival outcomes. In this work, we introduce a new framework that simultaneously performs patient clustering and shared-frailty survival modeling through a penalized likelihood approach. The proposed methodology adaptively learns a patient-to-patient similarity matrix via a modified version of spectral clustering, enabling cluster formation directly from estimated risk profiles while accounting for group membership. A simulation study highlights the proposed model's ability to recover latent clusters and to correctly estimate hazard parameters. We apply our method to a large cohort of heart-failure patients hospitalized with COVID-19 between 2020 and 2021 in the Lombardy region (Italy), identifying clinically meaningful subgroups characterized by distinct risk profiles and highlighting the role of respiratory comorbidities and hospital-level variability in shaping mortality outcomes. This framework provides a flexible and interpretable tool for risk-based patient stratification in hierarchical data settings.

Keywords: frailty models, clustering, penalization, heterogeneity, survival analysis

1 Introduction

In real-world time-to-event studies arising from clinical and epidemiological studies, identifying patients with similar risk trajectories enables the formation of subgroups, or clusters, characterized by shared likelihoods of experiencing key outcomes (e.g., death, relapse, treatment failure). Such stratification, relevant across a variety of application domains, in this setting can support clinical decision-making, facilitate disease subtyping, and advance precision-medicine initiatives (Collins and Varmus, 2015). Nonetheless, patients are often intrinsically dependent, as they may share group membership – such as being treated in the same hospital – inducing correlations in outcomes and additional variability that must be accounted for in the analysis, as they can influence patients' risk profiles in complex ways.

This scenario is common in healthcare applications and is exemplified by the project motivating of our work, *Enhance-Heart*, a study conducted in the Northern Italy regional district (Lombardy region), a region severely affected during the early stages of the COVID-19 pandemic due to an uncontrolled spread of SARS-CoV-2 infections (Alteri et al., 2021). This study focuses on patients with heart failure (HF) and aims, among other objectives, to investigate the complex relationship between patients' clinical profiles and COVID-19–related mortality, while also accounting for differences across hospitals. Indeed, recent research has highlighted a significant interplay between heart failure and COVID-19 during the pandemic, with pre-existing heart failure complicating COVID-19 management and prognosis, and potentially leading to complications (Bader et al., 2021; Adeghate et al., 2021). Thus, identifying latent subgroups of HF patients hospitalized for COVID-19 while assessing how respiratory pathologies and overall healthcare condition influence mortality risk across patient clusters, represents a key step in characterizing outcomes in a time-to-event framework. On the other hand, hospital-based resource availabilities, such as intensive care unit beds and general medicine/surgical beds, were found to be

significantly associated with incidence rate of COVID-19 deaths (Janke et al., 2021). Therefore, an accurate assessment of survival outcomes in this context requires modeling approaches capable of stratifying patient populations while accounting for differences across hospitals.

Traditional survival models often address heterogeneity arising from group-level dependence through a frailty term, that is a multiplicative factor in the hazard function, i.e. the risk of experiencing an event at a given time conditional on survival up to that point (Vaupel et al., 1979; Clayton, 1978). This term can be specified in different ways, either parametrically – most commonly using Gamma or Log-normal distributions – or semiparametrically (see, e.g., Abrahantes et al. (2007); Austin (2017); Balan and Putter (2020)). However, these models often assume homogeneity within groups and overlook latent similar patterns between patients induced by their risk of experiencing an event of interest at a given time. Conversely, unsupervised stratification methods can reveal hidden patient subgroups by leveraging similarities across individuals; however, their limited integration with time-to-event outcomes often leads to suboptimal predictive performance. For example, Ahlqvist et al. (2018) performed a cluster analysis on six standardized clinical measures, identifying replicable diabetes subgroups with distinct prognoses and complication profiles. Advancing toward survival-informed clustering, Tosado et al. (2020) proposed transforming the feature space according to each feature's association with survival (via martingale residuals) before applying clustering in this transformed space. Similarly, Chen et al. (2016) introduced a recursive partitioning algorithm that derives prognostic cancer staging systems by maximizing survival separation between groups of patients. Shifting the focus towards prediction, Chapfuwa et al. (2020) developed a Bayesian nonparametric framework that embeds subjects in a clustered latent space to enable accurate time-to-event predictions and subgroups with distinct risk profiles. Along related lines, Manduchi et al. (2021) presented a variational-autoencoder-based survival clustering approach that jointly leverages patient features and time-to-event outcomes, while Buginga and de Souza e Silva (2024) proposed a nonparametric mixture-of-experts model in which each expert captures a distinct subgroup's survival distribution, thereby uncovering heterogeneous risk patterns. To the best of our knowledge, relatively few statistical models address the joint problem of clustering patients and modeling time-to-event outcomes while incorporating hierarchical structure. Among these, the cluster-weighted model of Caldera et al. (2025) represents a more recent contribution, in which a shared frailty model is fitted within each cluster and covariates are assumed to follow parametric distributions different in each cluster, with the overall joint density factorized via cluster-specific mixing weights. While this approach is promising, its reliance on strong distributional assumptions for the covariates may restrict its flexibility.

To address this gap, we propose a novel model that extends parametric shared frailty models to account for group membership (e.g., patients treated in the same hospital) while simultaneously estimating both patient cluster membership – defined by similar risk trajectories – and hazard function parameters via penalized likelihood optimization, where the resulting solution is obtained through an iterative procedure. Patient stratification is achieved through a modified version of spectral clustering approach following the strategies proposed in Nie et al. (2014, 2016); Guo (2015), consisting in the adaptive estimation of a similarity matrix capturing patient-to-patient relationships representing the probability that patients are *neighbors*, based on their estimated hazard risk, combining frailty and covariate effects. At the end of the procedure, graph-theoretic principles ensure that the connected components of the resulting graph, represented by the learned similarity matrix, correspond exactly to the number of clusters.

Our model builds upon the framework proposed by Liu et al. (2020), originally developed for cancer subtyping, which integrates supervised graph-based clustering into the modeling of multi-omic tumor data through penalization terms embedded in a *partial* likelihood-based optimization scheme. Related extensions and applications of this approach have been explored in radiomics settings in Cavinato et al. (2021, 2023).

A key advantage of this framework is that patient clustering is embedded directly within the optimization procedure, so that cluster assignments are guided by underlying risk dynamics and reflect distinct event hazard profiles, making them interpretable in terms of event risk. Building on this framework, our contribution is twofold:

(i) we extend the model to account for group-level (hospital) dependence together with patient-level heterogeneity within a unified survival modeling framework;

(ii) we propose a penalized *full*-likelihood approach that embeds patient clustering directly within the hazard model, allowing similarity structures to be learned in a risk-driven manner and producing clusters that are interpretable in terms of event risk.

This paper is organized as follows. Section 2 outlines the proposed methodology, detailing the notation and the estimation procedure. In Section 3, we present a simulation study designed to evaluate the performance of the proposed methodology. In Section 4, we describe the *Enhance-Heart* administrative database that motivated our work and report the main results obtained with our method. Finally, Section 5 provides concluding remarks, discusses the strengths and limitations of the proposed approach, and outlines directions for future work.

The method, implemented through the software R (R Core Team (2022)), is openly accessible at https://github.com/alessandragni/HazClust.

2 Methodology

In this section, we outline the methodology following three consecutive steps: a recap on notation for parametric frailty models (Subsection 2.1), the model description (Section 2.2), the model optimization (Section 2.3) and some details related to the estimation procedure (Section 2.4).

2.1 Notation

Let the observed data be $\mathcal{Z} = \{(y_{gi}, \delta_{gi}, \boldsymbol{x}_{gi}) \mid i = 1, \dots, n_g,\ g = 1, \dots, M\}$, where units $i = 1, \dots, n_g$ are nested within groups $g = 1, \dots, M$, and the total number of units is $N = \sum_{g=1}^{M} n_g$. For each unit, $y_{gi} = \min(t_{gi}, C^*_{gi})$ denotes the observed time to event or censoring, whichever comes first, and $\delta_{gi} = \mathbb{I}(t_{gi} \leq C^*_{gi})$ indicates whether the event was observed ($\delta_{gi} = 1$) or censored ($\delta_{gi} = 0$). The covariate vector $\boldsymbol{x}_{gi} \in \mathbb{R}^{d\times 1}$ may include both group-level and unit-specific components. We assume independent and noninformative censoring (Munda and Legrand, 2014). To model the hazard rate of subject i in group g at time t, accounting for the presence of heterogeneity across groups, among the possible approaches (e.g, fixed effects, stratified), we consider the shared frailty model (Duchateau and Janssen, 2008)

$$h_{gi}(t \mid u_g) = h_0(t; \boldsymbol{\psi}) u_g \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta}) \tag{1}$$

where $\boldsymbol{\beta} \in \mathbb{R}^{d\times 1}$ is the vector of fixed-effect coefficients, the u_g's, $g = 1, \dots, M$, are the actual values of a random variable with probability density $f_U(\cdot \mid \theta)$, that is the frailty distribution, and $h_0(t; \boldsymbol{\psi})$ is the baseline hazard, assumed parametric, depending on the vector of parameters $\boldsymbol{\psi}$, and $'$ stands for the transpose. The cumulative hazard then follows as $H_{gi}(t \mid u_g) = u_g \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta}) H_0(t; \boldsymbol{\psi})$, with $H_0(t; \boldsymbol{\psi}) = \int_0^t h_0(s; \boldsymbol{\psi}) ds$. Following computations in Appendix A, the marginal loglikelihood can be expressed as

$$\ell(\boldsymbol{\psi}, \boldsymbol{\beta}, \theta, \mathcal{Z}) = \sum_{g=1}^{M} \left[\left[\sum_{i=1}^{n_g} \delta_{gi} \left(\log(h_0(y_{gi}; \boldsymbol{\psi})) + \boldsymbol{x}'_{gi}\boldsymbol{\beta} \right) \right] + \log \left[(-1)^{d_g} \mathcal{L}^{(d_g)} \left(\sum_{i=1}^{n_g} H_0(y_{gi}; \boldsymbol{\psi}) \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta}) \right) \right] \right] \tag{2}$$

where $d_g := \sum_{i=1}^{n_g} \delta_{gi}$ is the number of events in group g and $(-1)^k \mathcal{L}^{(k)}(s) := (-1)^k \frac{d^k}{ds^k} \mathcal{L}(s) = \mathbb{E}[U^k \exp(-Us)]$, being $\mathcal{L}(s)$ the Laplace transform, for $s \geq 0$ (Munda and Legrand, 2014; Klein, 1992).

Finally, it is known that the frailty term u_g can be predicted by $\hat{u}_g = \mathbb{E}\big(U \mid \mathcal{Z}_g; \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\beta}}, \hat{\theta}\big)$, being $\mathcal{Z}_g$ the observed data in group g and

$$\mathbb{E}\big(U \mid \mathcal{Z}_g; \boldsymbol{\beta}, \boldsymbol{\psi}, \theta\big) = -\frac{\mathcal{L}^{(d_g+1)}\Big(\sum_{i=1}^{n_g} H_0(y_{gi}; \boldsymbol{\psi}) \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta}) \Big)}{\mathcal{L}^{(d_g)}\Big(\sum_{i=1}^{n_g} H_0(y_{gi}; \boldsymbol{\psi}) \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta}) \Big)} \,. \tag{3}$$

2.2 Model

In a parametric setting, such as that addressed by (Munda et al., 2012), where both the frailty and the baseline hazard follow specified parametric distribution, the model parameters $\boldsymbol{\beta}, \boldsymbol{\psi}, \theta$ can be estimated by maximizing the loglikelihood in (2), or minimizing the negative loglikelihood, as follows

$$\min_{\boldsymbol{\beta}, \boldsymbol{\psi}, \theta} \quad -\ell(\boldsymbol{\psi}, \boldsymbol{\beta}, \theta, \mathcal{Z}) \,. \tag{4}$$

With the aim of performing clustering on the units while learning $\boldsymbol{\beta}, \boldsymbol{\psi}$ and θ, let $\mathbf{S} \in \mathbb{R}^{N\times N}$ be the affinity matrix whose entries $\{s_{gi,hj}\}$ represent the similarity between the unit i in group g and the unit j in group h. Specifically, we interpret each entry $\{s_{gi,hj}\}$ as the probability that unit i in group g has unit j in group h as a neighbor, imposing the constraint $\sum_{h=1}^{M} \sum_{j=1}^{n_h} s_{gi,hj} = 1$, $\forall g, i$ that ensures that the

connection probabilities for each unit sum up to one. By assuming that units have a higher probability of being neighbors if they are more similar, or equivalently have a lower distance $d_{gi,hj}(\boldsymbol{\psi}, \boldsymbol{\beta}, \theta, \mathcal{Z})$, we rewrite the optimization problem in Eq. (4) by further penalizing the loglikelihood for jointly estimating $\boldsymbol{\beta}, \boldsymbol{\psi}, \theta$ and $\mathbf{S}$:

$$\begin{aligned} \min_{\boldsymbol{\beta},\boldsymbol{\psi},\theta,\mathbf{S}} \quad & -\ell(\boldsymbol{\psi}, \boldsymbol{\beta}, \theta, \mathcal{Z}) + \gamma \sum_{g,h=1}^{M} \sum_{i=1}^{n_g} \sum_{j=1}^{n_h} \left(d_{gi,hj}(\boldsymbol{\beta}, \theta, \mathcal{Z})\, s_{gi,hj} + \mu\, s_{gi,hj}^2 \right) \\ \text{s.t.} \quad & \sum_{h=1}^{M} \sum_{j=1}^{n_h} s_{gi,hj} = 1, \quad \mathbf{s}_{gi} \geq 0, \quad \forall\, i = 1, \ldots, n_g,\ g = 1, \ldots, M \end{aligned} \tag{5}$$

where $\mathbf{s}_{gi} \in \mathbb{R}^{1\times N}$ is the row of $\mathbf{S}$ corresponding to unit i in group g, and the reported inequality is elementwise. The last two terms in Eq. (5) follow the literature on clustering with adaptive neighbors, see for instance Nie et al. (2014). These terms regulate the structure of neighborhood probabilities in $\mathbf{S}$: the first term encourages higher probabilities for nearby units, that, if alone leads to the limit case of 1 for the nearest neighbor and 0 elsewhere, while the quadratic term promotes smoother and more evenly distributed assignments, with the limit case assigning equal probability $1/N$ to all units (Nie et al., 2014). The hyperparameter γ controls the strength of the penalization; when $\gamma = 0$, no penalization is applied, and consequently no clustering is induced, as the penalty and its associated constraints vanish. In contrast, μ is a tunable parameter that modulates the influence of the similarity structure in the overall optimization. Its selection will be discussed in Sections 2.3 and 2.4.

As we aim to learn $\mathbf{S}$ based on the on the similarity between the risk of experiencing the event of interest for patient i in group g, compared to patient j in group h, that we quantify on the log-hazard scale, as $\log h_{gi}(t \mid u_g) = \log h_0(t; \boldsymbol{\psi}) + \log u_g + \boldsymbol{x}'_{gi}\boldsymbol{\beta}$ and $\log h_{hj}(t \mid u_h)$, where the $\log h_0(t; \boldsymbol{\psi})$ shared across units cancels out in pairwise comparisons. Thus, we define

$$d_{gi,hj}(\boldsymbol{\beta}, \theta, \mathcal{Z}) := \left\| (\boldsymbol{x}'_{gi}\boldsymbol{\beta} + \log u_g) - (\boldsymbol{x}'_{hj}\boldsymbol{\beta} + \log u_h) \right\|^2, \tag{6}$$

where $\|\cdot\|$ as the Euclidean distance, and u_g and u_h are the frailty terms defined in (3). The resulting distance encourages the similarity graph to reflect frailty-adjusted risk alignment.

To perform clustering into C clusters, we interpret $\mathbf{S} \in \mathbb{R}^{N\times N}$ as the adjacency matrix of a graph with the N units as nodes. In this setting, we define the Laplacian matrix $\mathbf{L_S} := \mathbf{D_S} - (\mathbf{S}' + \mathbf{S})/2$, where $\mathbf{D_S} \in \mathbb{R}^{N\times N}$ is the diagonal degree matrix, where the gi-th entry is $\sum_{hj}(s_{gi,hj} + s_{hj,gi})/2$. Following Mohar et al. (1991), the multiplicity C of the zero eigenvalue of $\mathbf{L_S}$ corresponds to the number of connected components in the graph with similarity matrix $\mathbf{S}$. Enforcing $rank(\mathbf{L_S}) = N - C$ thus promotes a partition into C clusters. Furthermore, as shown in Nie et al. (2014) and Fan (1949), this rank constraint can be handled equivalently and more easily by introducing a matrix $\mathbf{F} \in \mathbb{R}^{N\times C}$, given that

$$2Tr(\mathbf{F}'\mathbf{L_S}\mathbf{F}) = \sum_{g,h=1}^{M} \sum_{i=1}^{n_g} \sum_{j=1}^{n_h} \|\mathbf{f}_{gi} - \mathbf{f}_{hj}\|^2 s_{gi,hj}, \tag{7}$$

being $\mathbf{f}_{gi} \in \mathbb{R}^{C\times 1}$ the row of $\mathbf{F}$ corresponding to the belonging of unit i in cluster g.

Combining (2), (5) and (6), we get an optimization problem that reads

$$\begin{aligned} \min_{\boldsymbol{\beta},\boldsymbol{\psi},\theta,\mathbf{S},\mathbf{F}} \quad & -\left\{ \sum_{g=1}^{M} \left[\left[\sum_{i=1}^{n_g} \delta_{gi} \left(\log(h_0(y_{gi}; \boldsymbol{\psi})) + \boldsymbol{x}'_{gi}\boldsymbol{\beta} \right) \right] + \log \left[(-1)^{d_g} \mathcal{L}^{(d_g)} \left(\sum_{i=1}^{n_g} H_0(y_{gi}; \boldsymbol{\psi}) \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta}) \right) \right] \right] \right\} \\ & + \gamma \left\{ \sum_{g,h=1}^{M} \sum_{i=1}^{n_g} \sum_{j=1}^{n_h} \left[\left\| (\boldsymbol{x}'_{gi}\boldsymbol{\beta} + \log u_g) - (\boldsymbol{x}'_{hj}\boldsymbol{\beta} + \log u_h) \right\|^2 s_{gi,hj} + \mu\, s_{gi,hj}^2 \right] \right. \\ & \left. + 2\lambda Tr(\mathbf{F}'\mathbf{L_S}\mathbf{F}) \right\} \\ \text{s.t.} \quad & \sum_{h=1}^{M} \sum_{j=1}^{n_h} s_{gi,hj} = 1, \quad \mathbf{s}_{gi} \geq 0, \quad \forall\, i = 1, \ldots, n_g,\ g = 1, \ldots, M, \quad \mathbf{F}'\mathbf{F} = \mathbf{I},\ \mathbf{F} \in \mathbb{R}^{N\times C} \end{aligned} \tag{8}$$

where

$$\log u_g = \log \left[(-1)^{d_g+1} \mathcal{L}^{(d_g+1)} \left(\sum_{i=1}^{n_g} H_0(y_{gi}; \boldsymbol{\psi}) \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta}) \right) \right] - \log \left[(-1)^{d_g} \mathcal{L}^{(d_g)} \left(\sum_{i=1}^{n_g} H_0(y_{gi}; \boldsymbol{\psi}) \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta}) \right) \right],$$

which follows from (3) by noting that $-1 = (-1)^{d_g+1}/(-1)^{d_g}$ and applying a logarithmic property.

2.3 Optimization procedure

To solve the optimization problem in Eq. (8), we adopt a block coordinate descent approach: at each iteration, we update one block of variables while keeping the others fixed, until a convergence criterion is met (see Subsection 2.4 for details). The procedure alternates among the following three steps:

1. **Fix S, $\boldsymbol{\beta}, \boldsymbol{\psi}, \theta$ and update F**
   With $\mathbf{S}$, $\boldsymbol{\beta}$, $\boldsymbol{\psi}$, and θ fixed, the update of $\mathbf{F}$ reduces to the following spectral optimization problem (Nie et al., 2014):

$$\begin{aligned} \min_{\mathbf{F}} \quad & 2\lambda Tr(\mathbf{F}^T \mathbf{L_S} \mathbf{F}) \\ \text{s.t.} \quad & \mathbf{F}^T \mathbf{F} = \mathbf{I}, \mathbf{F} \in \mathbb{R}^{N \times C} \end{aligned}$$

   The factor 2λ does not affect the solution and can be omitted. The optimal $\mathbf{F}$ is obtained by taking the C eigenvectors of the Laplacian $\mathbf{L_S}$ associated with its C smallest eigenvalues. This step requires the number of clusters C and the current estimate of $\mathbf{S}$.

2. **Fix F, S and update $\boldsymbol{\beta}, \boldsymbol{\psi}, \theta$**
   With $\mathbf{F}$ and $\mathbf{S}$ fixed, the optimization over $(\boldsymbol{\beta}, \boldsymbol{\psi}, \theta)$ reduces to a survival modeling problem with an additional similarity-based regularization term. The optimization problem can be expressed as:

$$\min_{\boldsymbol{\beta}, \boldsymbol{\psi}, \theta} \quad -\ell(\boldsymbol{\beta}, \boldsymbol{w}, \theta, \mathcal{Z}) + \gamma \sum_{g,h=1}^{M} \sum_{i=1}^{n_g} \sum_{j=1}^{n_h} \left(\left\| \boldsymbol{x}'_{gi} \boldsymbol{\beta} + \log u_g - \boldsymbol{x}'_{hj} \boldsymbol{\beta} - \log u_h \right\|^2 \right) s_{gi,hj}$$

   We estimate the survival model parameters relying on maximum likelihood optimization with parametric frailty, following Munda et al. (2012), and extend this framework with the similarity-based penalty. This step requires γ, the current similarity matrix $\mathbf{S}$, the dataset and model specification.

3. **Fix F, $\boldsymbol{\beta}, \boldsymbol{\psi}, \theta$ and update S**
   With $\mathbf{F}$, $\boldsymbol{\beta}$, $\boldsymbol{\psi}$, and θ fixed, the update of $\mathbf{S}$ is obtained by solving

$$\begin{aligned} \min_{\mathbf{S}} \quad & \gamma \Bigg\{ \sum_{g,h=1}^{M} \sum_{i=1}^{n_g} \sum_{j=1}^{n_h} \left[\left\| \boldsymbol{x}'_{gi} \boldsymbol{\beta} + \log u_g - \boldsymbol{x}'_{hj} \boldsymbol{\beta} - \log u_h \right\|^2 s_{gi,hj} + \mu s_{gi,hj}^2 \right] \\ & + 2\lambda Tr(\mathbf{F}' \mathbf{L_S} \mathbf{F}) \Bigg\} \\ \text{s.t.} \quad & \sum_{h=1}^{M} \sum_{j=1}^{n_h} s_{gi,hj} = 1, \quad 0 \le \mathbf{s}_{gi} \le 1, \quad \forall\, i = 1, \ldots, n_g,\ g = 1, \ldots, M \end{aligned}$$

   Using relation (7) and the separability of the problem across units, this reduces to an independent optimization for each $\mathbf{s}_{gi}$:

$$\begin{aligned} \min_{\mathbf{s}_{gi}} \quad & \sum_{h=1}^{M} \sum_{j=1}^{n_h} \left(\left\| \boldsymbol{x}'_{gi} \boldsymbol{\beta} + \log u_g - \boldsymbol{x}'_{hj} \boldsymbol{\beta} - \log u_h \right\|^2 + \mu s_{gi,hj} + \lambda \| \mathbf{f}_{gi} - \mathbf{f}_{hj} \|^2 \right) s_{gi,hj} \\ \text{s.t.} \quad & \sum_{h=1}^{M} \sum_{j=1}^{n_h} s_{gi,hj} = 1, \quad \mathbf{s}_{gi} \ge 0 \end{aligned} \tag{9}$$

   As shown in Appendix B, this problem admits a closed-form solution for each row $\mathbf{s}_{gi}$:

$$s_{gi,hj} = \max \left\{ 0, \alpha_{gi} - w_{gi,hj} / (2\mu_{gi}) \right\}$$

   subject to $\mathbf{s}'_{gi} \mathbf{1} = 1$, where $\alpha_{gi} \ge 0$ and and μ_{gi} may differ across units. Here, $\boldsymbol{w}_{gi} \in \mathbb{R}^{N \times 1}$ and

$$w_{gi,hj} := \left\| \boldsymbol{x}'_{gi} \boldsymbol{\beta} + \log u_g - \boldsymbol{x}'_{hj} \boldsymbol{\beta} - \log u_h \right\|^2 + \lambda \| \mathbf{f}_{gi} - \mathbf{f}_{hj} \|^2.$$

   By enforcing sparsity on $\mathbf{s}_{gi}$, i.e., allowing only the k nearest neighbors to connect unit i in group g, following Nie et al. (2014); Guo (2015) we impose additional constraints that make the problem well-posed while also reducing computational cost. Let $\hat{\boldsymbol{w}}_{gi} = (\hat{w}_{gi,1}, \ldots, \hat{w}_{gi,N-1})$ denote the sorted distances (excluding the self-distance $w_{gi,gi}$). Then the optimal ordered similarity vector $\hat{\mathbf{s}}_{gi}$ is constrained to have exactly k nonzero entries:

(i) $\hat{s}_{gi,p} > 0$ for $p = 1, \ldots, k$, corresponding to the k nearest neighbors;

(ii) $\hat{s}_{gi,p} = 0$ for $p = k+1, \ldots, N$, including self-similarity.

Under these conditions, and the constraint $\mathbf{s}'_{gi}\mathbf{1} = 1$, as shown in Appendix C, we obtain

$$\alpha_{gi} = \frac{1}{k} + \frac{1}{2k\mu_{gi}} \sum_{p=1}^{k} \hat{w}_{gi,p} \quad \text{and} \quad \mu_{gi} = \frac{k}{2}\hat{w}_{gi,(k+1)} - \frac{1}{2}\sum_{p=1}^{k} \hat{w}_{gi,p}. \tag{10}$$

Finally, once all rows are updated, we set $\mu = \sum_{g,i} \mu_{gi}/N$.

This step requires λ, k, the current estimate of $\mathbf{F}$, the dataset and the model specification.

2.4 Estimation details

This section provides practical details and technicalities on the estimation and implementation of the proposed algorithm.

Initialization. If no prior information is provided, the similarity matrix $\mathbf{S}$ is initialized with zeros on the diagonal and $1/(N-1)$ for all off-diagonal entries, so each row sums to 1. An initial similarity matrix can alternatively be supplied via `S0`. The penalty parameter λ is set to $\lambda_0 = 1000$ by default, as it is advised in the literature to set it large enough (Nie et al., 2014), but may be specified by the user through the parameter `lambda0`. After the first update of $\mathbf{S}$, λ is reset to the value of μ obtained from the closed-form solution, and then adaptively adjusted: it is increased (multiplied by 1.2) if the number of connected components in $\mathbf{S}$ is smaller than C, or decreased (divided by 1.2) otherwise (Nie et al., 2017).

The regression and frailty parameters are initialized either based on user input via the `inip` and `iniFpar` options, or using default values as done in the `parfm` R package (Munda et al., 2012). As in this package, our software supports a range of parametric baseline hazard distributions (Weibull, exponential, Gompertz, log-normal, ...), as well as several frailty distributions (gamma, log-normal, positive stable, among others).

Convergence. Convergence is declared when the algorithm stabilizes both the similarity matrix and the model fit. Specifically, the procedure stops when all of the following are satisfied: (i) the mean absolute change in $\mathbf{S}$ is below `tolS`, indicating that the graph structure has stabilized; (ii) the improvement in log-likelihood is below `tolll`, showing that the statistical fit has plateaued; and (iii) the number of connected components in $\mathbf{S}$ equals the target C. These criteria are evaluated jointly, and the algorithm terminates when all are met or when a maximum of `maxit` iterations is reached. Additionally, the inner optimization in step 2, which riles on `parfm`, can be limited to `maxitparfm` iterations. Defaults are set to `tolS` $= 10^{-4}$, `tolll` $= 10^{-3}$, `maxit` $= 500$, `maxitparfm` $= 500$.

Choice of k and its relation to C. The number of nearest neighbors k controls the sparsity of the initial similarity matrix $\mathbf{S}$ and thereby the density of the corresponding graph. Indeed, each row contains exactly k nonzero entries, enforcing a fixed local neighborhood structure and excluding self-similarity (the diagonal of $\mathbf{S}$ is set to zero).

The choice of k strongly influences how readily the constructed graph can achieve the target of C connected compontents. If k is too small, the graph may fragment into more than C components, whereas a large k yields a dense graph that tends to collapse structure into fewer than C components. Our algorithm compensates for this interaction by adaptively tuning the penalty parameter λ to enforce convergence to the prescribed number C. However, in practice, k should be selected to guarantee basic connectivity without sacrificing sparsity, and needs to be carefully tuned depending on the application. A common heuristic is to let k grow moderately with the sample size.

Model selection and choice of C. To assess the quality of the clustering solution, we employ the silhouette coefficient (Kaufman and Rousseeuw, 2009). Given a unit i, let $a(i)$ denote its average distance to all other units in the same cluster, and let $b(i)$ be the minimum average distance from i to any other cluster. We recall that the silhouette value is defined as

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i),\, b(i)\}}, \quad s(i) \in [-1, 1],$$

with higher values indicating better separation. In our setting, distances for unit i with respect to units j are computed as $\sqrt{d_{gi,hj}(\boldsymbol{\beta}, \theta, \mathcal{Z})}$ in Eq. (6), as it incorporates the covariate effects and the estimated frailty contributions. Group indices (such as g for i and h for j) are omitted here since they play no role in the silhouette calculation. We then assess the overall clustering quality using the mean silhouette,

and select the optimal clustering configuration as the one with the maximum mean silhouette, thereby achieving the best separation with respect to the fitted survival model.

Software. The algorithm is implemented in R and is available in at the following repository https://github.com/alessandragni/HazClust. The core survival and frailty estimation routines rely on the `parfm` package (Munda et al., 2012), which provides maximum likelihood estimation for parametric frailty models. Our contribution extends `parfm` by embedding it within an iterative graph-based clustering framework.

3 Simulation Study

In this section, we present a simulation study to assess the performance of the proposed methodology. Our main focus is on evaluating how well the method recovers the underlying cluster structure across different levels of penalization. We also investigate the behavior of the parameter estimates under varying penalization strengths. Additionally, when the latent structure contains C clusters, we assess how accurately the method identifies the true clustering compared to alternative choices of C, with the mean silhouette providing a useful overall indicator of clustering quality.

3.1 Data generating process

We generate data by inducing the presence of $M = 10$ groups, each containing $n_g = 50$ ($g = 1, \dots, M$), yielding a total sample size of $N = 500$. To induce dependence among units within the same group, we introduce a group-level frailty u_g, drawn independently for each group from a Gamma distribution $u_g \sim \text{Gamma}(1/\theta, 1/\theta)$, with $\theta = 0.5$ so that $\mathbb{E}[u_g] = 1$ and $\text{Var}(u_g) = \theta$. Keeping the frailty variance small ensures that cluster-specific effects remain the dominant source of heterogeneity. Further elaboration of this aspect will be provided in the final Discussion.

Within each group g, every unit i is randomly assigned to one of the $C = 3$ latent clusters, denoted $z_{gi} \in \{1, 2, 3\}$. Each unit is associated with a covariate vector $\boldsymbol{x}_{gi}$ and a linear predictor $\eta_{gi} = \boldsymbol{x}'_{gi}\boldsymbol{\beta} + \log u_g$ where $\boldsymbol{\beta}$ is a vector of regression coefficients, so that, conditional on the cluster assignment z_{gi} and frailty u_g, the linear predictor is modeled as

$$\eta_{gi} \mid z_{gi} = c, u_g \sim \mathcal{N}(\mu_c, \sigma^2), \quad c = 1, 2, 3$$

where $[\mu_1, \mu_2, \mu_3] = [-8, 0, 8]$ and $\sigma = 1$. This specification generates clearly separated distributions of η_{gi} across clusters, maximizing the clustering structure as measured in Eq. (6). Finally, to preserve the intended simulation structure across clusters, we construct a single observed covariate x_{gi} as $x_{gi} = \big(\eta_{gi} - \log u_g\big)/\beta$, setting $\beta = \log(2)$.

Survival times are generated under a Weibull baseline hazard, parameterized by $\boldsymbol{\psi} = [\rho, \psi]$ through $H_0(t) = \xi^\rho t^\rho$ ($\rho, \xi > 0$), with $\rho = 2.5$ and $\xi = 0.01$. The corresponding inverse cumulative hazard is $H_0^{-1}(s) = s^{1/\rho}/\xi$. Event times are generated using inverse transform sampling: we draw $V_{gi} \sim \text{Uniform}(0, 1)$ and set $t_{gi} = H_0^{-1}\big(-\log(V_{gi})/\exp(\eta_{gi})\big)$.

Censoring is imposed using one of two possible mechanisms: (i) *administrative* censoring at a fixed time of 100, and (ii) *Gaussian (normal)* censoring with mean 130 and standard deviation 15, analogous to the default option of the `genfrail` function of the `frailtySurv` R package (Monaco et al., 2018).

Under this data generation mechanims, the Kaplan–Meier curves of each of the generated data highlight distinct patterns across the three clusters: one maintains high survival, another declines gradually, and a third drops sharply early on.

Following the data generating process just described, we set

$$\gamma \in \{0, 10^{-6}, 10^{-4}, 10^{-3}, 10^{-2}, 0.1, 0.2, 0.4\}.$$

The case $\gamma = 0$ corresponds to absence of penalization, so no clustering is performed, and just a parametric frailty model as in `parfm` is fitted. When $\gamma > 0$, we set $C \in \{2, 3, 4, 5\}$ and $k = \{20, 50\}$. For each setting, we replicate the simulation $B = 100$ times. In each replication, the same simulated dataset is used, while the input parameters of the model are varied to assess its behavior under different configurations. Furthermore, we set `tolS` $= 10^{-2}$, `tolll` $= 1$ and `maxit` $= 200$.

3.2 Results

We first focus on the simulations with $C = 3$ and $\gamma > 0$. For both $k = 20$ and $k = 50$, the method consistently identifies three clusters, with only minor differences in the γ thresholds at which performance begins to decline. Under *administrative* censoring, the method identifies the three clusters with 100% frequency for $\gamma \leq 0.1$. For larger γ, the identification rate decreases slightly, reaching a minimum of 95% at $\gamma = 0.4$. Overall, convergence was achieved after an average of 17.49 iterations. Under *normal* censoring, perfect identification is maintained for $\gamma \leq 10^{-3}$, after which the rate gradually slightly declines, reaching a minimum of 98% at $\gamma = 0.4$. In this second case, the algorithm required on average 24.49 iterations to converge.

Considering now only the cases in which three clusters were correctly identified, clustering performance is evaluated using accuracy and the adjusted Rand index (ARI), with the true cluster memberships of each unit i in group g known from the simulation design. Accuracy measures the proportion of observations whose predicted cluster labels match the true labels after optimal label alignment. ARI quantifies pairwise agreement corrected for chance, ranging in $[-1, 1]$, with 1 indicating perfect recovery, 0 random clustering, and negative values worse-than-chance agreement. Results are reported in Table 1 across different model configurations, that is, varying censoring mechanism, k, and γ.

Censoring	k	γ	**Accuracy**			**ARI**		
			Mean	Median	SD	Mean	Median	SD
(i) Administrative	20	10^{-6}	0.971	1.000	0.089	0.942	1.000	0.161
		10^{-4}	0.971	1.000	0.089	0.942	1.000	0.161
		10^{-3}	**0.975**	**1.000**	**0.081**	**0.950**	**1.000**	**0.147**
		10^{-2}	0.957	1.000	0.082	0.907	1.000	0.172
		0.1	0.950	1.000	0.083	0.888	1.000	0.179
		0.2	0.930	1.000	0.115	0.860	1.000	0.197
		0.4	0.892	0.928	0.141	0.788	0.819	0.242
	50	10^{-6}	**1.000**	**1.000**	**0.001**	**0.999**	**1.000**	**0.003**
		10^{-4}	**1.000**	**1.000**	**0.001**	**0.999**	**1.000**	**0.003**
		10^{-3}	**1.000**	**1.000**	**0.001**	**0.999**	**1.000**	**0.003**
		10^{-2}	1.000	1.000	0.001	0.999	1.000	0.004
		0.1	0.996	1.000	0.023	0.980	1.000	0.054
		0.2	0.945	0.998	0.120	0.897	0.994	0.199
		0.4	0.420	0.390	0.109	0.060	0.019	0.159
(ii) Normal	20	10^{-6}	0.902	0.998	0.142	0.810	0.994	0.246
		10^{-4}	0.904	0.998	0.138	0.810	0.994	0.251
		10^{-3}	0.891	0.970	0.141	0.788	0.918	0.243
		10^{-2}	0.896	0.998	0.149	0.807	0.994	0.239
		0.1	0.913	1.000	0.117	0.820	1.000	0.224
		0.2	**0.922**	0.998	**0.097**	**0.831**	0.994	**0.195**
		0.4	0.870	0.894	0.145	0.752	0.732	0.237
	50	10^{-6}	0.990	1.000	0.040	0.976	1.000	0.094
		10^{-4}	0.990	1.000	0.040	0.977	1.000	0.094
		10^{-3}	0.990	1.000	0.041	0.976	1.000	0.096
		10^{-2}	**0.998**	**1.000**	**0.016**	**0.994**	**1.000**	**0.034**
		0.1	0.994	1.000	0.027	0.986	1.000	0.060
		0.2	0.964	0.998	0.089	0.922	0.994	0.183
		0.4	0.417	0.386	0.102	0.056	0.015	0.148

Table 1: Empirical means, medians, and standard deviations of *accuracy* and *ARI* over $B = 100$ replications, evaluated under two censoring mechanisms and across varying values of k and γ, setting $C = 3$. The best estimates in each setting are highlighted in bold.

Clustering performance is generally very high, especially for larger k, with accuracy and ARI frequently approaching 1. Specifically, the median remains close to 1 across most settings, reflecting the skewed distribution of performance measures near the upper bound; the mean, in contrast, shows the gradual rise and the sharp drop more clearly: performance is slightly lower for very small γ, reaches its maximum at moderate γ of the chosen set, and then declines sharply once γ reaches 0.4, as penalization becomes much stronger. When accuracy and ARI are highest, their standard deviations are extremely small, indicating highly stable performance across replications.

Focusing still on the case $C = 3$, we now restrict attention to $k = 20$ to ease visualisation and reduce the number of scenarios presented. In this setting, we examine the estimates of the unknown parameters in the shared frailty model, $[\theta, \rho, \xi, \beta]$ (we recall that in Section 2, we named $\boldsymbol{\psi} = [\rho, \xi]$ as the Weibull distribution is parameterized by two parameters). For completeness, we also include $\gamma = 0$, which allows us to visualise the parameter estimates obtained in the absence of penalisation. Figures 1 and 2 display boxplots of the estimated parameters across different values of γ, for administrative and normal censoring, respectively. In these boxplots, the true parameter values are indicated by a dashed horizontal line, allowing a visual assessment of bias. As expected, the bias in the estimates generally increases with γ, represented on the x-axis. In particular, the parameter θ in plot (a), representing the variance of the frailty, shows a slight underestimation even in the unpenalized case ($\gamma = 0$, first boxplot), a pattern that is consistent with previous studies relying on the implementation of `parfm` package (Munda et al., 2012). Moreover, most parameters show relatively stable behaviour in terms of variability across different γ values, indicating that penalization has a limited effect on their spread. The only exception is ξ, the second parameter of the Weibull, which appears to be the most affected, with its variability increasing noticeably as γ increases.

To evaluate how penalization affects the quality of the parameter estimates, Table D.1 reports their variance (Var) and the mean squared error normalized by variance (MSE/Var) across all values of γ and both censoring schemes. Appendix D provides a detailed interpretation of these results.

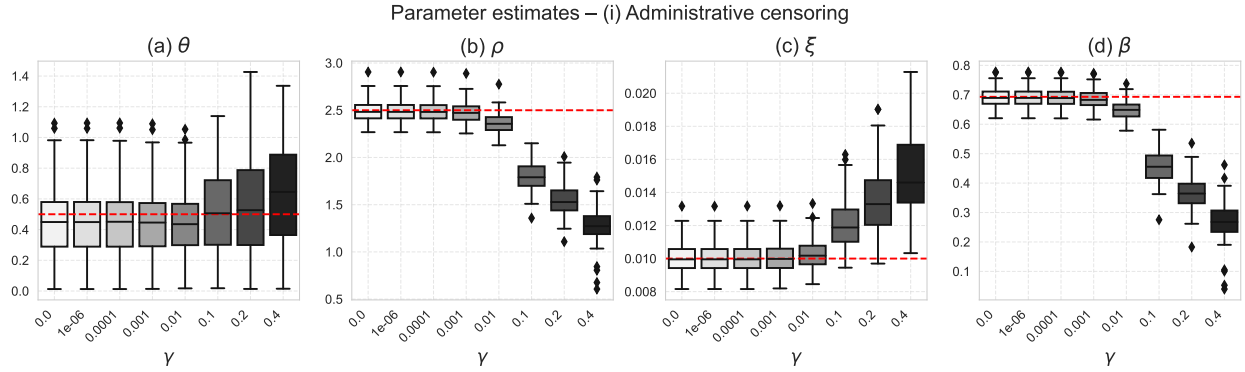

Figure 1: Boxplots of parameter estimates $(\theta, \rho, \xi, \beta)$ across penalization levels γ (x-axis) under *administrative* censoring for $k = 20$ and $C = 3$. The true parameter values are shown as a piecewise-constant horizontal line.

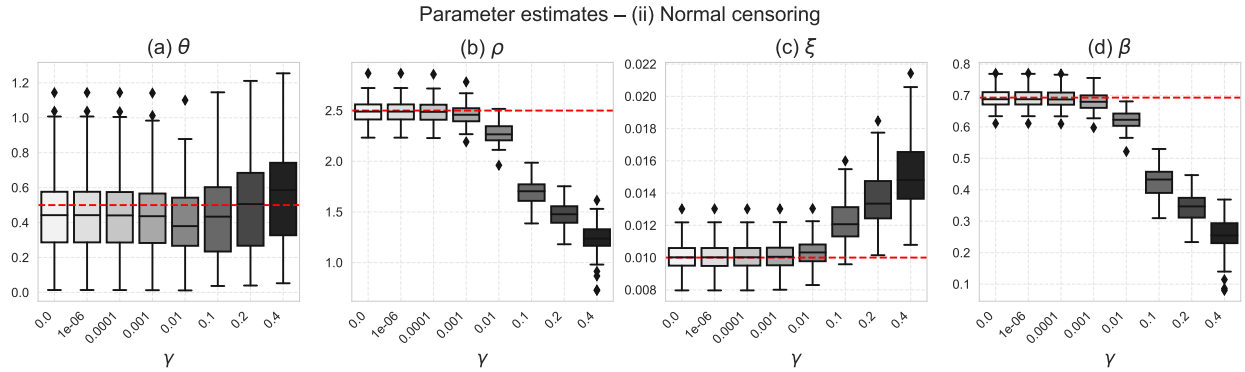

Figure 2: Boxplots of parameter estimates $(\theta, \rho, \xi, \beta)$ across penalization levels γ (x-axis) under *normal* censoring for $k = 20$ and $C = 3$. The true parameter values are shown as a piecewise-constant horizontal line.

We now turn to analysing the results obtained varying C. In Table 2, we analyse the empirical distribution of the mean silhouette values for varying values of C under the two censoring mechanisms,

and a specific choice of $k = 20$ and $\gamma = 10^{-4}$. The results indicate that $C = 3$ yields the highest average and median mean silhouette values in both scenarios, suggesting that this choice provides the best overall cluster separation. Although the silhouette values for $C = 3$ exhibit slightly greater variability, the overall separation remains substantially stronger. Moreover, convergence to the selected number of clusters within the prescribed number of iterations is always reached at $C = 3$ across all replications, while it is not the case for all the other choices.

Censoring	C	**Mean silhouette**			**Convergence**
		Mean	Median	SD	%
(i) Administrative	2	0.660	0.670	0.046	67
	3	**0.792**	**0.832**	**0.116**	**100**
	4	0.661	0.718	0.121	100
	5	0.566	0.612	0.113	100
(ii) Normal	2	0.618	0.649	0.082	93
	3	**0.670**	**0.789**	**0.182**	**100**
	4	0.541	0.577	0.173	99
	5	0.466	0.484	0.145	100

Table 2: Summary of the empirical distribution of the mean silhouette values over $B = 100$ replications under two censoring mechanisms, for $k = 20$ and $\gamma = 10^{-4}$. For each value of C, the table reports the mean, median, standard deviation, and the proportion of runs that converged to the selected C within `maxit` iterations. The estimates related to $C = 3$ are highlighted in bold.

4 Case Study

This section applies the proposed methodology to the *Enhance-Heart* dataset, a large administrative dataset linked with the Covid-19 registry[1] of Lombardy Region (RL in the following).

4.1 Cohort selection

Sourcing from the administrative dataset related to the *Enhance-Heart* project, we select those patients that between *January 1, 2018*, and *December 31, 2020* were diagnosed with HF, that is, hospitalized or visited ER under DRG code 127 ('Heart Failure and Shock') according to the RL MS-DRG v40 system, thus including both primary/secondary diagnoses of HF (ICD-9-CM: 428.xx) and related conditions (e.g. ICD-9-CM: 402.01, 402.11, 402.91). Among these, we retain those hospitalized for COVID-19 according to a binary flag provided in the database, between *31 January 2020* (start of the national COVID-19 state of emergency in Italy) and *18 June 2021* (implementation of the Green Pass regulations governing reopening). For those patients, outcomes are assessed over a 90-day follow-up period; death within this window is considered the event, while patients surviving beyond or dying later are regarded as censored (55.25% of the sample).

The analysis is restricted to hospitals with a minimum of 50 patients hospitalized, as smaller hospitals may exhibit different organizational conditions and characteristics, and to ensure more robust results for the frailty-related estimates. We further exclude patients who required hospital transfers for medical purposes or who had multiple episodes of infection. The final cohort includes $N = 3086$ HF patients hospitalized for COVID-19 in $J = 32$ different hospitals in the RL.

The time-to-event outcome is defined by the pair (`Time`, `Status`), where `Time` denotes days until death within the 90-day observation window or censoring, and `Status` indicates event occurrence through a binary variable. Patient's covariates include `Age` at admission, `Gender`, and Modified Multisource Comorbidity Score (`ModMCS`), a measure of overall comorbidity burden unrelated to the primary diagnosis and a proxy for general health status, with higher values reflecting greater comorbidity. It was introduced by Corrao et al. (2017) and adapted by Caldera et al. (2025) to exclude respiratory diseases relevant to COVID-19. Indeed, we have considered separately the variable `Resp` defining it as a binary indicator that equals 1 when more than one respiratory condition - among chronic obstructive pulmonary disease, bronchitis, pneumonia, or respiratory failure - is present. A detailed description and summary statistics of the study variables are reported in Table 3.

[1] https://www.dati.lombardia.it/

Table 3: List, description and summary statistics of selected variables in the RL database.

Variable	Description	Type	Summary statistics
`Status`	Binary indicator for death, equal to 1 if the event occurs within the observation period, 0 otherwise.	Categorical	44.75% for 1 (death) and 55.25% for 0 (alive).
`Time`	Either the survival time (in days) or the observation period, depending on `Status`.	Continuous	mean = 19.20, sd = 18.04, median = 12, [Q1; Q3] = [6; 26], [min; max] = [1; 89].
`Age`	Patient's age at hospital admission.	Continuous	mean = 81.36, sd = 8.27, median = 82.0, [Q1; Q3] = [76; 88], [min; max] = [60; 106]
`Gender`	Patient's biological sex (male/female).	Categorical	55.83% for male, 44.17% for female.
`ModMCS`	Modified Multisource-Comorbidity Score (Caldera et al., 2025; Corrao et al., 2017) quantifying cumulative burden of other diseases.	Continuous	mean = 11.47, sd = 7.77, median = 10, [Q1; Q3] = [6; 16], [min; max] = [0; 57]
`Resp`	Count of respiratory diseases among *chronic obstructive pulmonary disease*, *bronchitis*, *pneumonia*, and *respiratory failure*.	Continuous	mean = 0.97, sd = 1.18, median = 1, [Q1; Q3] = [0; 2], [min; max] = [0; 4]

4.2 Model setup and results

We fit our model under multiple configurations and identify the optimal one as that with the maximum mean silhouette value, as described in Section 2.4 (Kaufman and Rousseeuw, 2009). Specifically, we consider two choices for the frailty distribution (Gamma and Inverse Gaussian) and three options for the baseline hazard distribution (Weibull and Exponential). For the penalization parameters, we explore $\gamma = \{10^{-8}, 10^{-5}, 10^{-2}, 10^{-1}\}$ and $k \in \{50, 100\}$. Finally, we evaluate models with $C \in \{2, 3, 4\}$. We set `maxit` $= 200$, `tolS` $= 10^{-2}$, `tolll` $= 1$ and leave all the other parameters with default settings. The mean silhouette values for all the above configurations are reported in Table 4.

We select the model with the frailty distributed as Inverse Gaussian, the baseline as Weibull, $\gamma = 10^{-1}$, $k = 150$ and $C = 3$ as the optimal clustering configuration, as it yields the highest mean silhouette score equal to 0.609 and therefore the best separation with respect to the fitted survival model. With the chosen model we obtain the following regression coefficients: $\hat{\beta}_{\texttt{Age}} = 0.0025$, $\hat{\beta}_{\texttt{Gender=male}} = 0.0244$, $\hat{\beta}_{\texttt{ModMCS}} = 0.0013$, and $\hat{\beta}_{\texttt{Resp}} = 0.0033$. The corresponding p-values are all < 0.05, indicating statistically significant effects. The estimated frailty variance is $\theta = 0.573$, while the Weibull shape and scale parameters were $\rho = 0.618$ and $\lambda = \psi^{\rho} = 1.126$, respectively.

The model reaches convergence in 23 iterations. The three obtained clusters – corresponding to disconnected components in the patients graph – contain, respectively, 1077, 1054 and 955 patients, revealing subgroups with differing risk profiles. Specifically, the first (and largest) cluster includes older patients (mean age 83.5 years) with the highest modified multisource-comorbidity score (average `ModMCS` equal to 12.6) while moderate respiratory disease burden (average `Resp` equal to 0.99), predominantly males, and has a mean follow-up time of 0.147. Cluster 2 comprises slightly younger patients (mean age 81.8 years) with intermediate `ModMCS` (11.8) and the highest respiratory disease burden (1.03), also male-predominant, with a mean follow-up of 0.158. Cluster 3, the smallest one, contains the youngest patients (mean age 78.4 years) with the lowest `ModMCS` (9.8) and respiratory disease burden (0.87), more females than males, and had the longest mean follow-up (0.176).

Cluster membership varied substantially across hospitals. The median proportion of observations belonging to the dominant cluster was 75.2% (interquartile range 66.9–95.5%), indicating moderate within-hospital heterogeneity overall. Thirteen hospitals (40.6%) showed a strong predominance of a single cluster, with at least 80% of observations assigned to that cluster, while five hospitals (15.6%) were entirely associated with one cluster. In contrast, only three hospitals (9.4%) exhibited a mixed cluster composition, with no cluster accounting for more than 60% of observations.

The wide range in cluster patterns – from hospitals dominated by a single cluster to those with a more mixed composition – underscores substantial within-hospital heterogeneity and supports using an approach specifically designed to handle this variability rather than assuming uniformity. This heterogeneity is further reflected in the estimated θ, indicating meaningful variation in hospital-level effects that our method accounts for.

Parametric Distributions		Penalization Param.		C		
u_g	$H_0(t;\boldsymbol{\psi})$	γ	k	2	3	4
Gamma	Weibull	10^{-8}	50	0.544	0.470	0.320
			150	0.535	0.526	0.277
		10^{-5}	50	0.542	0.379	0.416
			150	0.526	0.496	0.293
		10^{-2}	50	0.526	0.508	0.292
			150	0.518	0.464	0.461
		10^{-1}	50	0.474	0.363	0.252
			150	0.591	**0.608**	0.033
Gamma	Exponential	10^{-8}	50	0.542	0.376	0.243
			150	0.548	0.473	0.196
		10^{-5}	50	0.545	0.290	0.277
			150	0.533	0.514	0.092
		10^{-2}	50	0.538	0.495	0.440
			150	0.525	0.404	0.379
		10^{-1}	50	0.524	0.126	0.441
			150	0.493	0.132	0.402
Inverse Gaussian	Weibull	10^{-8}	50	0.545	0.512	0.399
			150	0.545	0.498	0.419
		10^{-5}	50	0.525	0.462	0.170
			150	0.527	-0.244	0.187
		10^{-2}	50	0.532	0.491	0.426
			150	0.528	0.420	0.412
		10^{-1}	50	0.487	0.481	0.224
			150	0.594	**0.609**	0.001
Inverse Gaussian	Exponential	10^{-8}	50	0.544	0.452	0.434
			150	0.546	0.485	0.473
		10^{-5}	50	0.543	0.286	0.472
			150	0.525	0.305	0.417
		10^{-2}	50	-0.228	0.472	0.514
			150	0.533	0.437	0.427
		10^{-1}	50	0.536	0.046	0.209
			150	0.502	0.141	0.126

Table 4: Mean silhouette values across model configurations, varying frailty and baseline hazard distributions, penalization parameters γ and k, and number of clusters C.

5 Discussion

Motivated by the need to identify latent patient subgroups in hierarchical time-to-event data, in this work we propose a framework that jointly performs patient risk-based clustering and survival modeling while accounting for group-level dependence through a parametric shared frailty model, ensuring that cluster assignments are informed by underlying survival dynamics rather than by covariate similarity alone. From a methodological perspective, our approach addresses a gap in the existing literature by jointly modeling survival outcomes, latent patient subgroups, and group-level dependence within a single likelihood-based framework, allowing patient stratification to emerge from estimated risk profiles while properly accounting for within-hospital correlation.

The simulation study demonstrates that the proposed methodology reliably recovers latent cluster structures across a broad range of penalization settings, provided that the frailty variance is kept small so that cluster-specific effects remain the dominant source of heterogeneity. This choice facilitates identifiability by preventing individual-level random effects from obscuring systematic differences in underlying risk trajectories. Under these conditions, the method also achieves accurate estimation of the parametric frailty model when penalization is appropriately tuned using the silhouette coefficient as the selection criterion for both optimal clustering and parameter estimation.

When applied to the Enhance-Heart cohort of heart-failure patients hospitalized for COVID-19, the method identifies three clusters of patients with distinct risk profiles and clinically interpretable differences in age, comorbidity burden, respiratory conditions, and survival patterns. The clusters were obtained by selecting the parameter configuration and parametric distributions that maximize the silhouette coefficient, ensuring both optimal separation and cohesion. These clusters correspond to disconnected components in the patient graph and reveal clinically meaningful subgroups: the first cluster captures older patients with higher comorbidity but moderate respiratory burden, the second includes slightly younger patients with the highest respiratory burden, and the third comprises the youngest patients with the lowest comorbidity and respiratory burden, highlighting how the methodology can uncover nuanced heterogeneity in event risk patterns.

Beyond patient-level stratification, our results highlight substantial heterogeneity across hospitals. Cluster composition varied markedly between hospitals, with many institutions showing a strong predominance of a single cluster, while a smaller subset exhibited more heterogeneous patient mixes. Furthermore, the non-negligible value of the frailty parameter suggests that latent institutional factors – such as resource availability, clinical protocols, or contextual pressures during the COVID-19 pandemic – may influence survival outcomes beyond what is captured by observed covariates and cluster membership. This finding underscores the importance of modeling hierarchical dependence explicitly and supports the combined use of clustering and frailty terms as complementary tools for disentangling multiple sources of heterogeneity.

Some limitations warrant consideration. First, the construction of the similarity graph relies on the choice of the neighborhood size parameter k, which influences graph sparsity and cluster formation. Although adaptive tuning of the penalization parameter λ embedded in the algorithm mitigates this issue in practice, the relationship between k and the target number of clusters C remains an area for further investigation. Furthermore, each unit is forced to connect to its k nearest neighbors regardless of whether these lie near the cluster core or at its boundary. This limitation has been noted in the spectral clustering literature (Cai et al., 2022), where pruning weak connections and modifying the optimization term through power-function asymptotics have been proposed as potential remedies. However, such approaches introduce additional tuning parameters, including a maximum neighborhood size, and did not yield performance improvements in our setting.

In addition, in the current framework, a common baseline hazard is assumed across clusters, which is simplified in the penalization distance, while heterogeneity is captured through covariates and frailty terms. Future work will aim to extend the model to allow cluster-specific baseline hazards, enabling a more nuanced characterization of risk dynamics across patient subgroups. This extension would require reconsidering the frailty structure, as different clusters would have distinct frailties, going beyond the scopes of the present work. Additional developments could explore alternative similarity measures, as well as extensions to semi-parametric or fully nonparametric baseline hazards, further increasing the model's flexibility.

Competing interests

No competing interest is declared.

Data availability

The participants of this study did not give written consent for their data to be shared publicly, so due to the sensitive nature of the research and regulations applied from RL sharing terms, the full supporting data is not available.

Acknowledgements

This work is part of the *Enhance-Heart* project: Efficacy evaluatioN of the therapeutic-care patHways, of the heAlthcare providers effects aNd of the risk stratifiCation in patiEnts suffering from HEART failure. The authors acknowledge the 'Unità Organizzativa Osservatorio Epidemiologico Regionale' and ARIA S.p.A for providing data and technological support, and the financial support from the Department of Mathematics of Politecnico di Milano through the grant 'Dipartimento di Eccellenza 2023-2027' by MUR, Italy.

A Derivation of the marginal likelihood

The *conditional likelihood* for the g-th cluster is expressed as

$$L_g^c(\boldsymbol{\psi}, \boldsymbol{\beta} \mid u_g, \mathcal{Z}) = \prod_{i=1}^{n_g} \left(h_0(y_{gi}) u_g \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta})\right)^{\delta_{gi}} \cdot \exp\left(-H_0(y_{gi}) u_g \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta})\right)$$

and the *marginal likelihood* is obtained by integrating out the frailty distribution over its support S_U:

$$\begin{aligned}
L_g(\boldsymbol{\psi}, \boldsymbol{\beta}, \theta, \mathcal{Z}) &= \int_{S_U} L_g^c(\boldsymbol{\psi}, \boldsymbol{\beta} \mid u_g, \mathcal{Z}) \cdot f_U(u_g \mid \theta) du_g \\
&= \int_{S_U} \prod_{i=1}^{n_g} \left(h_0(y_{gi}) u_g \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta})\right)^{\delta_{gi}} \cdot \exp\left(-H_0(y_{gi}) u_g \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta})\right) \cdot f_U(u_g \mid \theta) du_g \\
&= \prod_{i=1}^{n_g} \left(h_0(y_{gi}) \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta})\right)^{\delta_{gi}} \int_{S_U} \prod_{i=1}^{n_g} (u_g)^{\delta_{gi}} \cdot \exp\left(-H_0(y_{gi}) u_g \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta})\right) \cdot f_U(u_g) du_g \\
&= \prod_{i=1}^{n_g} \left(h_0(y_{gi}) \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta})\right)^{\delta_{gi}} \int_{S_U} u_g^{\sum_{i=1}^{n_g} \delta_{gi}} \cdot \exp\left(-\sum_{i=1}^{n_g} H_0(y_{gi}) \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta}) u_g\right) \cdot f_U(u_g) du_g \\
&= \prod_{i=1}^{n_g} \left(h_0(y_{gi}) \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta})\right)^{\delta_{gi}} (-1)^{d_g} \mathcal{L}^{(d_g)}\left(\sum_{i=1}^{n_g} H_0(y_{gi}) \exp(\boldsymbol{x}'_{gi}\boldsymbol{\beta})\right)
\end{aligned}$$

where in the last equation we used $(-1)^k \mathcal{L}^{(k)}(s) := (-1)^k \frac{d^k}{ds^k}\mathcal{L}(s) := \mathbb{E}[U^k \exp(-Us)]$, being $\mathcal{L}(s)$, for $s \geq 0$, the Laplace transform (Munda and Legrand, 2014; Klein, 1992). By taking the logarithm and summing over the clusters, we obtain Eq. (2).

B Derivation of closed-form solution of (9)

Given that $\sum_{h,j} \mu s_{gi,hj}^2 + w_{gi,hj} s_{gi,hj} = \mu \|\mathbf{s}_{gi}\|^2 + \boldsymbol{w}'_{gi}\mathbf{s}_{gi} = \mu \|\mathbf{s}_{gi} + \frac{1}{2\mu}\boldsymbol{w}_{gi}\|^2 - \frac{1}{4\mu}\|\boldsymbol{w}_{gi}\|^2$, problem (9) can be rewritten as

$$\min_{\mathbf{s}'_{gi}\mathbf{1}=1, \mathbf{s}_{gi} \geq 0} \quad \left\|\mathbf{s}_{gi} + \frac{1}{2\mu}\boldsymbol{w}_{gi}\right\|^2$$

where the last term $\frac{1}{4\mu}\|\boldsymbol{w}_{gi}\|^2$ and the multiplicative μ were removed as the former does not depend on $\mathbf{s}_{gi}$ and the latter can be removed from the optimization problem. As we said that the problem is solved independently for each unit i in group g, in this section from now on we denote μ as μ_{gi}. This problem can be solved through the KKT conditions. Indeed, the Lagrangian function can be written as

$$\mathcal{L} = \left\|\mathbf{s}_{gi} + \frac{1}{2\mu_{gi}}\boldsymbol{w}_{gi}\right\|^2 - \alpha_{gi}(\mathbf{s}'_{gi}\mathbf{1} - 1) - \boldsymbol{r}'\mathbf{s}_{gi}$$

where $\alpha_{gi} \geq 0$ and $\boldsymbol{r} \geq 0$ are Lagrangian multipliers. According to the KKT condition, we have the following solution

$$s_{gi,hj} = \max\left\{0, \alpha_{gi} - \frac{w_{gi,hj}}{2\mu_{gi}}\right\} \tag{B1}$$

subject to $\mathbf{s}'_{gi}\mathbf{1} = 1$.

C Derivation of α_{gi} and μ_{gi} in (10)

The optimal solution is given by B1, subject to the constraint $\mathbf{s}'_{gi}\mathbf{1} = \sum_{h=1}^{N} s_{gi,hj} = 1$.

Let $\hat{w}_{gi,1} \leq \hat{w}_{gi,2} \leq \cdots \leq \hat{w}_{gi,N-1}$ denote the sorted distances (excluding the self-distance). Imposing k-nearest-neighbor sparsity yields

$$\begin{aligned} \hat{s}_{gi,p} &> 0, \quad p = 1, \ldots, k, \\ \hat{s}_{gi,p} &= 0, \quad p = k+1, \ldots, N. \end{aligned}$$

For the k nonzero entries,

$$\hat{s}_{gi,p} = \alpha_{gi} - \frac{\hat{w}_{gi,p}}{2\mu_{gi}}, \qquad p = 1, \ldots, k.$$

For the $(k+1)$-th neighbor, the sparsity condition implies

$$\alpha_{gi} - \frac{\hat{w}_{gi,k+1}}{2\mu_{gi}} \leq 0.$$

At optimality this constraint is active, giving

$$\alpha_{gi} = \frac{\hat{w}_{gi,k+1}}{2\mu_{gi}}.$$

Enforcing the simplex constraint over the k nonzero entries yields

$$\sum_{p=1}^{k} \hat{s}_{gi,p} = \sum_{p=1}^{k} \left(\alpha_{gi} - \frac{\hat{w}_{gi,p}}{2\mu_{gi}}\right) = 1,$$

$$k\alpha_{gi} - \frac{1}{2\mu_{gi}} \sum_{p=1}^{k} \hat{w}_{gi,p} = 1.$$

Solving for α_{gi} gives

$$\alpha_{gi} = \frac{1}{k} + \frac{1}{2k\mu_{gi}} \sum_{p=1}^{k} \hat{w}_{gi,p}. \tag{C1}$$

Combining (C) and (C1), we obtain

$$\frac{\hat{w}_{gi,k+1}}{2\mu_{gi}} = \frac{1}{k} + \frac{1}{2k\mu_{gi}} \sum_{p=1}^{k} \hat{w}_{gi,p},$$

$$2\mu_{gi} = k\hat{w}_{gi,k+1} - \sum_{p=1}^{k} \hat{w}_{gi,p}.$$

Therefore,

$$\mu_{gi} = \frac{k}{2}\hat{w}_{gi,k+1} - \frac{1}{2}\sum_{p=1}^{k} \hat{w}_{gi,p}.$$

D Parameter estimates: variability and estimation error

In this Appendix, we interpret Table D.1. The table reports the normalized mean squared error and variance of the parameter estimates for different penalization levels γ under both administrative and normal censoring. The estimators remain stable for very small penalties ($\gamma \leq 10^{-4}$), with MSE/Var values close to one across all parameters, indicating that estimation error is almost entirely driven by variance—an observation also confirmed by Figures 1 and 2. As γ increases into the range 10^{-3}–10^{-2}, bias begins to emerge, most noticeably for ρ and β, where MSE/Var rises despite relatively modest changes in variance. For larger penalties ($\gamma \geq 0.1$), estimation accuracy deteriorates sharply, with substantial inflation of MSE/Var, particularly for ρ and β. This pattern is observed under both censoring mechanisms but is consistently more pronounced when censoring occurs according to a normal distribution.

These findings should be interpreted jointly with Table 1, where an intermediate penalization level emerges as optimal, reflecting a trade-off between shrinkage and estimation error.

Censoring	γ	θ		ρ		ψ		β	
		MSE/Var	Var	MSE/Var	Var	MSE/Var	Var	MSE/Var	Var
(i) Administrative	0	1.022	0.051	1.006	0.013	1.001	0.000	1.007	0.001
	10^{-6}	1.022	0.051	1.006	0.013	1.001	0.000	1.007	0.001
	10^{-4}	1.023	0.051	1.008	0.013	1.001	0.000	1.010	0.001
	10^{-3}	1.025	0.050	1.042	0.012	1.003	0.000	1.056	0.001
	10^{-2}	1.046	0.046	2.633	0.011	1.077	0.000	3.190	0.001
	0.1	1.000	0.069	20.407	0.025	3.036	0.000	20.207	0.003
	0.2	1.047	0.091	35.186	0.027	4.269	0.000	33.404	0.003
	0.4	1.205	0.096	44.550	0.034	5.471	0.000	43.941	0.004
(ii) Normal	0	1.022	0.052	1.004	0.012	1.002	0.000	1.007	0.001
	10^{-6}	1.022	0.052	1.004	0.012	1.002	0.000	1.007	0.001
	10^{-4}	1.024	0.052	1.009	0.012	1.003	0.000	1.013	0.001
	10^{-3}	1.036	0.050	1.131	0.011	1.007	0.000	1.169	0.001
	10^{-2}	1.170	0.041	5.808	0.010	1.137	0.000	6.310	0.001
	0.1	1.077	0.054	46.782	0.014	3.922	0.000	39.592	0.002
	0.2	1.001	0.068	77.124	0.014	5.371	0.000	64.944	0.002
	0.4	1.058	0.078	74.959	0.022	6.449	0.000	71.004	0.003

Table D.1: Normalized mean squared error (MSE/Var) and variance (Var) of the parameter estimates for $k = 20$ and $C = 3$. For each parameter $p \in \theta, \rho, \xi, \beta$, we define $\text{MSE}(\hat{p}) = \mathbb{E}[\|\hat{p} - p\|_2^2]$ and $\text{Var}(\hat{p}) = \mathbb{E}[\|\hat{p} - \mathbb{E}[\hat{p}]\|_2^2]$, with p denoting the true parameter introduced in Section 3.1. The table reports results under administrative and normal censoring for all considered penalization levels γ.

References

Abrahantes, J. C., C. Legrand, T. Burzykowski, P. Janssen, V. Ducrocq, and L. Duchateau (2007). Comparison of different estimation procedures for proportional hazards model with random effects. *Computational statistics & data analysis 51*(8), 3913–3930.

Adeghate, E. A., N. Eid, and J. Singh (2021). Mechanisms of covid-19-induced heart failure: a short review. *Heart failure reviews 26*(2), 363–369.

Ahlqvist, E., P. Storm, A. Käräjämäki, M. Martinell, M. Dorkhan, A. Carlsson, P. Vikman, R. B. Prasad, D. M. Aly, P. Almgren, et al. (2018). Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *The lancet Diabetes & endocrinology 6*(5), 361–369.

Alteri, C., V. Cento, A. Piralla, V. Costabile, M. Tallarita, L. Colagrossi, S. Renica, F. Giardina, F. Novazzi, S. Gaiarsa, et al. (2021). Genomic epidemiology of sars-cov-2 reveals multiple lineages and early spread of sars-cov-2 infections in lombardy, italy. *Nature communications 12*(1), 434.

Austin, P. C. (2017). A tutorial on multilevel survival analysis: methods, models and applications. *International Statistical Review 85*(2), 185–203.

Bader, F., Y. Manla, B. Atallah, and R. C. Starling (2021). Heart failure and covid-19. *Heart failure reviews 26*(1), 1–10.

Balan, T. A. and H. Putter (2020). A tutorial on frailty models. *Statistical methods in medical research 29*(11), 3424–3454.

Buginga, G. and E. de Souza e Silva (2024). Clustering survival data using a mixture of non-parametric experts. *arXiv preprint arXiv:2405.15934*.

Cai, Y., J. Z. Huang, and J. Yin (2022). A new method to build the adaptive k-nearest neighbors similarity graph matrix for spectral clustering. *Neurocomputing 493*, 191–203.

Caldera, L., A. Cappozzo, C. Masci, M. Forlani, B. Antonelli, O. Leoni, A. M. Paganoni, and F. Ieva (2025). Cluster-weighted modeling of lifetime hierarchical data for profiling covid-19 heart failure patients. *arXiv preprint arXiv:2507.12230*.

Caldera, L., C. Masci, A. Cappozzo, M. Forlani, B. Antonelli, O. Leoni, and F. Ieva (2025, 04). Uncovering mortality patterns and hospital effects in covid-19 heart failure patients: a novel multilevel logistic cluster-weighted modeling approach. *Biometrics 81*(2), ujaf046.

Cavinato, L., N. Gozzi, M. Sollini, C. Carlo-Stella, A. Chiti, and F. Ieva (2021). Recurrence-specific supervised graph clustering for subtyping hodgkin lymphoma radiomic phenotypes. In *2021 43rd annual international conference of the IEEE engineering in medicine & biology society (EMBC)*, pp. 2155–2158. IEEE.

Cavinato, L., N. Gozzi, M. Sollini, M. Kirienko, C. Carlo-Stella, C. Rusconi, A. Chiti, and F. Ieva (2023). Perspective transfer model building via imaging-based rules extraction from retrospective cancer subtyping in hodgkin lymphoma. *Authorea Preprints*.

Chapfuwa, P., C. Li, N. Mehta, L. Carin, and R. Henao (2020). Survival cluster analysis. In *Proceedings of the ACM Conference on Health, Inference, and Learning*, pp. 60–68.

Chen, D., H. Wang, L. Sheng, M. T. Hueman, D. E. Henson, A. M. Schwartz, and J. A. Patel (2016). An algorithm for creating prognostic systems for cancer. *Journal of medical systems 40*(7), 160.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika 65*(1), 141–151.

Collins, F. S. and H. Varmus (2015). A new initiative on precision medicine. *New England journal of medicine 372*(9), 793–795.

Corrao, G., F. Rea, M. Di Martino, R. De Palma, S. Scondotto, D. Fusco, A. Lallo, L. M. B. Belotti, M. Ferrante, S. P. Addario, et al. (2017). Developing and validating a novel multisource comorbidity score from administrative data: a large population-based cohort study from italy. *BMJ open 7*(12), e019503.

Duchateau, L. and P. Janssen (2008). *The frailty model*. Springer.

Fan, K. (1949). On a theorem of weyl concerning eigenvalues of linear transformations i. *Proceedings of the National Academy of Sciences 35*(11), 652–655.

Guo, X. (2015). Robust subspace segmentation by simultaneously learning data representations and their affinity matrix. In *IJCAI*, pp. 3547–3553.

Janke, A. T., H. Mei, C. Rothenberg, R. D. Becher, Z. Lin, and A. K. Venkatesh (2021). Analysis of hospital resource availability and covid-19 mortality across the united states. *Journal of hospital medicine 16*(4), 211–214.

Kaufman, L. and P. J. Rousseeuw (2009). *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.

Klein, J. P. (1992). Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics*, 795–806.

Liu, C., W. Cao, S. Wu, W. Shen, D. Jiang, Z. Yu, and H.-S. Wong (2020). Supervised graph clustering for cancer subtyping based on survival analysis and integration of multi-omic tumor data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics 19*(2), 1193–1202.

Manduchi, L., R. Marcinkevičs, M. C. Massi, T. Weikert, A. Sauter, V. Gotta, T. Müller, F. Vasella, M. C. Neidert, M. Pfister, et al. (2021). A deep variational approach to clustering survival data. *arXiv preprint arXiv:2106.05763*.

Mohar, B., Y. Alavi, G. Chartrand, and O. Oellermann (1991). The laplacian spectrum of graphs. *Graph theory, combinatorics, and applications 2*(871-898), 12.

Monaco, J. V., M. Gorfine, and L. Hsu (2018). General semiparametric shared frailty model: estimation and simulation with frailtysurv. *Journal of statistical software 86*, 1–42.

Munda, M. and C. Legrand (2014). Adjusting for centre heterogeneity in multicentre clinical trials with a time–to–event outcome. *Pharmaceutical statistics 13*(2), 145–152.

Munda, M., F. Rotolo, and C. Legrand (2012). parfm: Parametric frailty models in r. *Journal of Statistical Software 51*, 1–20.

Nie, F., G. Cai, and X. Li (2017). Multi-view clustering and semi-supervised classification with adaptive neighbours. In *Proceedings of the AAAI conference on artificial intelligence*, Volume 31.

Nie, F., X. Wang, and H. Huang (2014). Clustering and projected clustering with adaptive neighbors. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 977–986.

Nie, F., X. Wang, M. Jordan, and H. Huang (2016). The constrained laplacian rank algorithm for graph-based clustering. In *Proceedings of the AAAI conference on artificial intelligence*, Volume 30.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Tosado, J., L. Zdilar, H. Elhalawani, B. Elgohari, D. M. Vock, G. E. Marai, C. Fuller, A. S. Mohamed, and G. Canahuate (2020). Clustering of largely right-censored oropharyngeal head and neck cancer patients for discriminative groupings to improve outcome prediction. *Scientific reports 10*(1), 3811.

Vaupel, J. W., K. G. Manton, and E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography 16*(3), 439–454.

MOX Technical Reports, last issues

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

03/2026 Mapelli, A.; Carini, L.; Ieva, F.; Sommariva, S.
A neighbour selection approach for identifying differential networks in conditional functional graphical models

02/2026 Antonietti, P.F.; Beirao da Veiga, L.; Botti, M.; Harnist, A.; Vacca, G.; Verani, M.
Virtual Element methods for non-Newtonian shear-thickening fluid flow problems

01/2026 Iapaolo V.; Vergani, A.M.; Cavinato, L.; Ieva, F.
Multi-view learning and omics integration: a unified perspective with applications to healthcare

Iapaolo, V.; Vergani, A.; Cavinato, L.; Ieva, F.
Multi-view learning and omics integration: a unified perspective with applications to healthcare

82/2025 Varetti, E.; Torzoni, M.; Tezzele, M.; Manzoni, A.
Adaptive digital twins for predictive decision-making: Online Bayesian learning of transition dynamics

81/2025 Leimer Saglio, C. B.; Pagani, S.; Antonietti, P. F.
A massively parallel non-overlapping Schwarz preconditioner for PolyDG methods in brain electrophysiology

79/2025 Zacchei, F.; Conti, P.; Frangi, A.; Manzoni, A.
Multi-Fidelity Delayed Acceptance: hierarchical MCMC sampling for Bayesian inverse problems combining multiple solvers through deep neural networks

78/2025 Botti, M.; Mascotto, L.
Trace inequalities for piecewise W^1,p functions over general polytopic meshes

Botti, M.; Mascotto, L.
Trace inequalities for piecewise W^{1,p} functions over general polytopic meshes

77/2025 Bonetti, S.; Botti, M.; Vega, P.
A robust fully-mixed finite element method with skew-symmetry penalization for low-frequency poroelasticity