



HAL
open science

Multivariate joint model under competing risks to predict death of hospitalized patients for SARS-CoV-2 infection

Alexandra Lavalley-morelle, Nathan Peiffer-Smadja, Simon B Gressens, Bérénice Souhail, Alexandre Lahens, Agathe Bounhiol, François-xavier Lescure, France Mentré, Jimmy Mullaert

► To cite this version:

Alexandra Lavalley-morelle, Nathan Peiffer-Smadja, Simon B Gressens, Bérénice Souhail, Alexandre Lahens, et al.. Multivariate joint model under competing risks to predict death of hospitalized patients for SARS-CoV-2 infection. *Biometrical Journal*, 2023, 66 (1), 10.1002/bimj.202300049 . hal-04454216

HAL Id: hal-04454216

<https://hal.science/hal-04454216v1>

Submitted on 13 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH ARTICLE

Multivariate joint model under competing risks to predict death of hospitalized patients for SARS-CoV-2 infection

Alexandra Lavalley-Morelle¹  | Nathan Peiffer-Smadja^{1,2} | Simon B. Gressens²  |
Bérénice Souhail² | Alexandre Lahens² | Agathe Bounhiol² |
François-Xavier Lescure^{1,2} | France Mentré^{1,3} | Jimmy Mullaert^{1,3}

¹Université Paris Cité, INSERM, IAME, Paris, France

²Department of Infectious and Tropical Diseases, AP-HP, Bichat-Claude Bernard University Hospital, Paris, France

³Department of Epidemiology, Biostatistics and Clinical Research, AP-HP, Bichat-Claude Bernard University Hospital, Paris, France

Correspondence

Alexandra Lavalley-Morelle, Université Paris Cité, INSERM, IAME, F-75018 Paris, France.

Email:

alexandra.lavalley-morelle@inserm.fr



This article has earned an open data badge “**Reproducible Research**” for making publicly available the code necessary to reproduce the reported results. The results reported in this article were reproduced partially due to data confidentiality issues.

Abstract

During the coronavirus disease 2019 (COVID-19) pandemic, several clinical prognostic scores have been proposed and evaluated in hospitalized patients, relying on variables available at admission. However, capturing data collected from the longitudinal follow-up of patients during hospitalization may improve prediction accuracy of a clinical outcome. To answer this question, 327 patients diagnosed with COVID-19 and hospitalized in an academic French hospital between January and July 2020 are included in the analysis. Up to 59 biomarkers were measured from the patient admission to the time to death or discharge from hospital. We consider a joint model with multiple linear or nonlinear mixed-effects models for biomarkers evolution, and a competing risks model involving sub-distribution hazard functions for the risks of death and discharge. The links are modeled by shared random effects, and the selection of the biomarkers is mainly based on the significance of the link between the longitudinal and survival parts. Three biomarkers are retained: the blood neutrophil counts, the arterial pH, and the C-reactive protein. The predictive performances of the model are evaluated with the time-dependent area under the curve (AUC) for different landmark and horizon times, and compared with those obtained from a baseline model that considers only information available at admission. The joint modeling approach helps to improve predictions when sufficient information is available. For landmark 6 days and horizon of 30 days, we obtain AUC [95% CI] 0.73 [0.65, 0.81] and 0.81 [0.73, 0.89] for the baseline and joint model, respectively ($p = 0.04$). Statistical inference is validated through a simulation study.

KEYWORDS

biomarkers, competing risks, COVID-19, dynamic predictions, multivariate joint model

1 | INTRODUCTION

The pandemic of coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), rapidly infected millions of people around the world, causing saturation of intensive care units (ICUs) and emergency departments. In this context, personalized predictions of the survival of hospitalized patients can be useful to guide therapeutic management (e.g., escalation or limitation of care) and to forecast hospital needs (beds, staff, etc.).

Since 2020, many prognostic scores (Knight et al., 2020; Liang et al., 2020; Myrstad et al., 2020) using information available at hospital admission (e.g., from medical history, clinical presentation, or biological measurements) have been developed to quantify the risk of in-hospital mortality for COVID-19 patients (Tjendra et al., 2020; Zhang et al., 2020; Zhu et al., 2020). These scores ignore the biological evolution during hospitalization, and attempt to use this information remain scarce. Some studies quantified the prognostic effect of an early change in biomarker measurements (e.g., between 2 days; Lavillegrand et al., 2021; Li et al., 2021; Mueller et al., 2020). However, when biomarkers evolve over time, incorporating the full follow-up in a statistical model may be useful to better predict disease prognostic. This can be achieved by jointly estimating a (possibly nonlinear) longitudinal mixed model describing the biomarkers evolution at individual level, and a survival model describing hazard evolution. These models are usually linked by shared random effects that model the association between the individual biomarkers evolution and their translation into hazard (Rizopoulos, 2011). Joint modeling also allows to derive individual dynamic predictions and recent examples of application are numerous (Desmée et al., 2017; Keroui et al., 2020; Lavalley-Morelle et al., 2022).

Nowadays, most hospitals are equipped with laboratory information systems that routinely gather all results of biological analyses prescribed by the physicians. During a hospital stay, consecutive clinical observations and biological results represent a huge amount of information that can be used in a joint model to provide individual dynamic predictions. Of note, when studying in-hospital death as a time-to-event outcome, patient discharge should be taken into account as a competing event. Shifting the modeling from one single biomarker to potentially hundreds could certainly improve prediction accuracy, but also comes with statistical issues. First, Shen et al. (Shen and Li, 2021) well described computational and identifiability issues due to the high number of random effects when multiple correlated longitudinal models are jointly estimated with a survival model. This probably explains why published models are mostly limited to longitudinal models with two biomarkers (Andrinopoulou et al., 2017; Long and Mills, 2018; Rajeswaran et al., 2018), and highlights the need of methods for variable selection in this context. Recently, a study (Tong-Minh et al., 2022) evaluated the follow-up of four immune biomarkers (C-reactive protein (CRP), procalcitonin, interleukin-6, and soluble urokinase-type plasminogen activator receptor) to predict the risk of in-hospital death in COVID-19 patients. This study provided promising results, but the involved biomarkers were selected based on clinical considerations and no statistical selection procedure is described. In a high-dimensional setting where biomarkers correspond to gene expression levels, Liu et al. (2019) proposed a factor analysis model to summarize the co-evolution of gene expressions. Again, this approach makes the estimation computationally tractable but does not perform marker selection since the loading matrix is full. In order to efficiently leverage information from hospital information systems, it appears that variable selection represents a second methodological challenge. While variable selection procedures are known for standard regression models, applications in joint models in the presence of competing risks are, to our knowledge, lacking.

The objective of the present work is twofold. Taking advantage on real data of patients hospitalized for SARS-CoV-2 infection, the first objective is to propose an original selection strategy to build a multivariate joint model with a subset of biomarkers most associated with the risk of in-hospital death and discharge from hospital. The second objective is to compare the predictive performances of such a model against a model that only consider baseline information to show the added value of the proposed strategy. This article is organized as follows. Section 2 introduces the analyzed data; Section 3 presents methodological aspects about uni- and multivariate joint modeling; Section 4 details the methodology to perform dynamic predictions and the performances assessment; Section 5 shows the results; and Section 6 provides a simulation study to support the methodology. Finally, Section 7 is devoted to the discussion.

2 | DATA

The study is based on real-life hospital data, obtained from $N = 327$ patients hospitalized in the Department of Infectious and Tropical Diseases in an academic hospital (Bichat, France) during the first wave of the COVID-19 pandemic (January–July 2020). The project has been approved by the local ethic committee (IRB number 00006477). A manual data

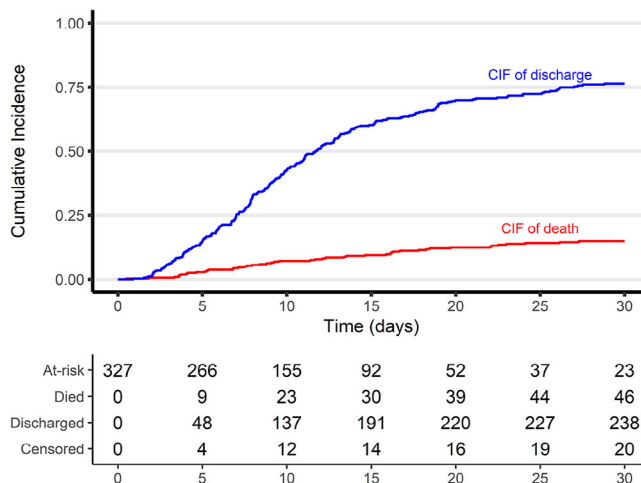


FIGURE 1 Cumulative incidence functions (CIF) for in-hospital death (red) and hospital discharge (blue). The number of events and at-risk patients is given at the bottom.

collection allowed to collect clinical variables relative to the patients: baseline characteristics (age, gender, etc.). The primary outcome is the time until in-hospital death within 30 days. Discharge is considered as a competing event. Figure 1 displays the cumulative incidence functions for the in-hospital death and the discharge from hospital. Thirty days after admission, about 14% of the patient died, while 72% were discharged. The censored observations correspond to hospital transfer during the stay or the end of the observation study (at day 30).

An automated extraction from the hospital data warehouse allowed to obtain all results of 59 biological exams prescribed during the hospitalization, up to day 30 or to the time to death or discharge from hospital. The availability and the frequency of the biomarker measurements differ between patients. These biomarkers were classified into eight categories: complete blood count, coagulation, pulmonary functions, urine samples, kidney functions/cellular lysis, liver/pancreatic functions, markers of inflammation, and cardiac markers. Table S1 shows the complete list description of these biomarkers.

In this study, we consider a baseline score adapted from the 4C score (Knight et al., 2020). This score was developed in 2020 to predict the in-hospital mortality for COVID-19 patients and combines eight variables (age, sex at birth, number of comorbidities, respiratory rate, peripheral oxygen saturation on air room, Glasgow coma scale, urea, and CRP at admission) with different weights (see Table S2). However, since the collected data do not include respiratory rate, peripheral oxygen saturation on room air, and Glasgow coma scale, we set to 0 the corresponding weights. Indeed, because the patients are hospitalized in a medical ward and not in an ICU, these items are likely to be closed to 0. For baseline urea and CRP values, we used the first available result within 48 h. The remaining missing values are imputed with the most frequently encountered value knowing the other characteristics of the score (age, sex, and number of comorbidities). Actually, it affects few individuals (4% and 8% of the patients for the urea and CRP, respectively). Table 1 details the score components of the analyzed patients. The median [Q1–Q3] score was 6 [4–9] at admission.

3 | BUILDING OF A JOINT MODEL WITH MULTIPLE BIOMARKERS

The main idea of this building process is to model each biomarker individually with a joint model, select the biomarkers properly fitted, and among them, keep the biomarkers whose evolution is significantly associated with the risk of in-hospital death. Then, the multivariate analysis is performed including the selected biomarkers.

3.1 | Univariate modeling

3.1.1 | General univariate model

Let T the variable describing the time-to-event distribution, C the noninformative censoring distribution, and δ the event indicator (1 for in-hospital death, 2 for discharge, and 0 for censoring). The couple (\tilde{T}, δ) is observed with $\tilde{T} = \min(T, C)$.

TABLE 1 Baseline characteristics (components of the baseline score) for the 327 patients of the data.

Number of patients	327
Age—number (%)	
<50	73 (22)
[50–59[74 (23)
[60–69[71 (22)
[70–79[62 (19)
≥80	47 (14)
Gender—male—number (%)	198 (61)
Comorbidities—number (%)	
0	139 (43)
1	93 (28)
≥2	95 (29)
Urea (mmol/L)—med [Q1–Q3]	5.6 [4.1–8.1]
NA—number (%)	2 (0.6)
1st measurement >48 h—number (%)	11 (3)
CRP (mg/L)—med [Q1–Q3]	67.5 [30.3–120.8]
NA—number (%)	9 (3)
1st measurement >48 h—number (%)	15 (5)
Score—med [Q1–Q3]	6 [4–9]

Abbreviation: NA: Not available, CRP: C-reactive protein.

We consider the subdistribution hazard framework to model the competing risks. In that case, we define $T_e = T \times \mathbb{1}_{\delta=e} + \infty \times \mathbb{1}_{\delta \neq e}$ for $e = \{1, 2\}$.

Let N be the total number of analyzed patients and K the number of biomarkers. Let N_k be the number of subjects who have at least one measurement for biomarker k and $\{y_{i1k}, \dots, y_{in_{ik}k}\}$ the vector of longitudinal observations of subject i (for $i = 1, \dots, N_k$) for this biomarker. Hence, observation y_{ijk} denotes the j th measurement of patient i for biomarker k at time t_{ijk} (for $j = 0, \dots, n_{ik}$). Let m_k be the function describing the structural model of biomarker k . Let also λ_{1ik} and λ_{2ik} be the instantaneous risks of in-hospital death and discharge, respectively, associated to individual i and biomarker k . We model each biomarker k , with a joint model defined as follows:

$$\begin{aligned}
 y_{ijk} &= m_k(t_{ijk}, \psi_{ik}) + g[m_k(t_{ijk}, \psi_{ik}), \sigma_k] \epsilon_{ij} \\
 \lambda_{1ik}(t|\psi_{ik}, \text{Score}_i) &= h_{1k} \exp[\alpha_{1k} \times (m_k(t, \psi_{ik}) - \text{med}_k) + \beta_{1k} \cdot \text{Score}_i] \\
 \lambda_{2ik}(t|\psi_{ik}, \text{Score}_i) &= h_{2k} \exp[\alpha_{2k} \times (m_k(t, \psi_{ik}) - \text{med}_k) + \beta_{2k} \cdot \text{Score}_i],
 \end{aligned} \tag{1}$$

where $\psi_{ik} = f(\mu_k, \eta_{ik})$ denotes the individual parameters associated to biomarker k expressed as a function of fixed effects, noted μ_k , and individual random effects, noted η_{ik} . The random effects are assumed to be normally distributed with mean 0 and variance-covariance matrix $\Omega_k = \text{diag}(\omega_k^2)$, and independent of the residual Gaussian error, noted ϵ_{ij} , of mean 0 and variance 1. σ_k denotes the vector of the error model parameters. Score_i denotes the baseline score associated to patient i . h_{1k} and h_{2k} are the baseline constant hazards for the risk of in-hospital death and the risk of discharge, respectively. α_{1k} and α_{2k} are the coefficients that link the current predicted value of the biomarker k with the instantaneous risks of events. We considered only a linear link between the longitudinal process and subdistribution risks. The median value of all observations for biomarker k is denoted as med_k and appears in the model in order to avoid numerical issues during estimation. Finally, β_{1k} and β_{2k} are the coefficients associated to the baseline score effects on both instantaneous risks. We emphasize that under a joint modeling approach, it is not necessary to know the value of the biomarker after the occurrence of the competing event since only the predictions given the structural model are linked with both hazards.

The structural model m_k describing the marker observations can be (1) a linear mixed-effects model or (2) a nonlinear mixed-effects model, which is more flexible for describing nonmonotonic trend, but also less parsimonious. In the case of

linear joint models, the function m_k is defined as follows:

$$m_k(t_{ijk}, \psi_{ik}) = b_{0ik} + b_{1ik} \times t_{ijk} \quad (2)$$

and for nonlinear joint models, the function m writes:

$$m_k(t_{ijk}, \psi_{ik}) = b_{0ik} + a_{ik} \times \left(\exp \left(b_{1ik} \times (t_{ijk} - t_{lag_{ik}}) \right) - \exp \left(b_{2ik} \times (t_{ijk} - t_{lag_{ik}}) \right) \right). \quad (3)$$

We have

$$\begin{cases} b_{0ik} &= \mu_{0k} + \eta_{0ik} \\ b_{1ik} &= \mu_{1k} + \eta_{1ik} \\ b_{2ik} &= \mu_{2k} + \eta_{2ik} \\ a_{ik} &= \mu_{ak} \exp(\eta_{aik}) \\ t_{lag_{ik}} &= \mu_{lag_k} + \eta_{t_{lag_{ik}}} \end{cases}$$

For a joint model associated to biomarker k , the joint likelihood writes:

$$L(\theta_k) = \prod_{i=1}^N \int_{\eta_{ik}} \left[\left(\prod_{j=0}^{n_{ik}} p(y_{ijk} | \eta_{ik}, Score_i; \theta_k) \right) p(\tilde{T}_i, \delta_i | \eta_{ik}; \theta_k) p(\eta_{ik}; \theta_k) \right] d\eta_{ik}, \quad (4)$$

where $\theta_k = \{\mu_k, \Omega_k, \sigma_k, h_{1k}, h_{2k}, \alpha_{1k}, \alpha_{2k}, \beta_{1k}, \beta_{2k}\}$ is the vector of parameters to estimate. $p(y_{ijk} | \eta_{ik}, Score_i; \theta_k)$ is the density function of longitudinal process associated to biomarker k , $p(\eta_{ik}; \theta_k)$ the density function of random effects and $p(t, \delta | \eta_{ik}; \theta_k)$ is the density function of the survival process given by (Jeong & Fine, 2006):

$$\begin{aligned} p(t, \delta | \eta_{ik}; \theta_k) &= [\lambda_{1ik}(t | \psi_{ik}, Score_i; \theta_k) \times (1 - F_{1ik}(t | \psi_{ik}, Score_i; \theta_k))]^{\delta=1} \\ &\quad \times [\lambda_{2ik}(t | \psi_{ik}, Score_i; \theta_k) \times (1 - F_{2ik}(t | \psi_{ik}, Score_i; \theta_k))]^{\delta=2} \\ &\quad \times (1 - F_{1ik}(t | \psi_{ik}, W_i; \theta_k) - F_{2ik}(t | \psi_{ik}, Score_i; \theta_k))^{\delta=0}, \end{aligned} \quad (5)$$

where, for $e = \{1, 2\}$:

$F_{eik}(t | \psi_{ik}, Score_i; \theta_k) = 1 - \exp(-\int_0^t \lambda_{eik}(u | \psi_{ik}, Score_i; \theta_k) du)$ is the individual cumulative incidence function for event e at time t .

The vector of population parameters θ can be estimated by maximization of the likelihood. However, to the best of our knowledge, no available R package can handle this likelihood expression. Therefore, we used the SAEM algorithm (Delyon et al., 1999) implemented on Monolix software version 2018R2, as in Lavalley-Morelle et al. (2022). Its variance-covariance matrix Σ_k is estimated by inversion of the Fisher information matrix (FIM), computed with a stochastic approximation using the Louis method (Louis, 1982). To ensure better estimation properties, the number of SAEM algorithm chains is fixed at 10 and the maximum number of iterations for both exploratory phase and stochastic approximation is set at 1000, keeping the default auto-stop criteria when convergence is reached. A previous work (Lavalley-Morelle et al., 2022) showed that Monolix software is able to provide unbiased and accurate estimates of parameters of a competing risk joint model in this context.

3.1.2 | Selection of biomarkers after univariate modeling

Linear submodel evaluation

We started by modeling each biomarker individually with the linear joint model for sake of simplicity. The initial values for the algorithm are set to the estimates obtained by modeling a “no-link” model (i.e., with the link coefficients α_{1k} and α_{2k} set to 0). A combined error is considered at this step (i.e., $g[m_k(t_{ijk}, \psi_{ik}), \sigma_k] = \sigma_{ak} + \sigma_{bk} \times m_k(t_{ijk}, \psi_{ik})$). After fitting $K = 59$ linear joint models, we obtain $\hat{\theta}_k$ the estimate of θ_k .

Error model selection

The following approach was used to select the error model associated to biomarker k : if $\hat{\sigma}_{bk} \leq 10^{-4}$, an additive error is considered for the biomarker ($g[m_k(t_{ijk}, \psi_{ik}), \sigma_k] = \sigma_{ak}$), if $\hat{\sigma}_{ak} \leq 10^{-4}$, a proportional error is considered for the biomarker ($g[m_k(t_{ijk}, \psi_{ik}), \sigma_k] = \sigma_{bk} \times m_k(t_{ijk}, \psi_{ik})$). Otherwise, the error model which minimizes the Bayesian Information Criteria (BIC) is considered. The models are then refitted with the correct error function, and we define:

$$er\hat{r}_k = \begin{cases} \frac{\hat{\sigma}_{ak}}{med_k} & \text{if the biomarker } k \text{ has an additive error} \\ \hat{\sigma}_{bk} & \text{if the biomarker } k \text{ has a proportional error} \\ \frac{\hat{\sigma}_{ak}}{med_k} + \hat{\sigma}_{bk} & \text{if the biomarker } k \text{ has a combined error.} \end{cases}$$

Fit criteria

Biomarkers with a small measurement error ($er\hat{r}_k \leq 25\%$) and with all relative standard errors (RSE) relative to the longitudinal parameters under 100% are kept with the linear joint model for the rest of the analysis. Biomarkers with a small measurement error ($er\hat{r}_k \leq 25\%$) but with at least one RSE of a longitudinal parameter over 100% are excluded from the analysis. Finally, biomarkers with a larger measurement error ($er\hat{r}_k > 25\%$) are evaluated using a nonlinear longitudinal submodel.

Nonlinear submodel

After the linear fit, some biomarkers were tested with the nonlinear submodel. First, due to potential identifiability issues, $\mu_{t_{lagk}}$ and $\omega_{t_{lagk}}$ are fixed to 0 if the RSE associated to parameter t_{lag} is over 100%. In fact, t_{lag} is an optional parameter that give more flexibility to the structural function, however, it can be easily removed to simplify the model. Then, same strategy is conducted for the selection and exclusion of the biomarkers but by being more flexible on the threshold for the measurement error. Hence, biomarkers with a measurement error $er\hat{r}_k \leq 50\%$ and with all RSE of longitudinal parameters under 100% are kept with the nonlinear joint model for the rest of the analysis. Biomarkers with a measurement error $er\hat{r}_k > 50\%$ or with at least one RSE of a longitudinal parameter over 100%, are excluded from the analysis.

Survival submodel evaluation

For each biomarker selected either with a linear or a nonlinear longitudinal submodel, the last step was to evaluate the survival part of the model. Two criteria are considered. The first involves the good quality of survival parameter estimation. The biomarkers for which the RSE of survival parameters are below 100% are selected for the next. The others are excluded from the analysis. The second criteria involves the link between the longitudinal process and the instantaneous risk of death. We want to select biomarkers whose longitudinal evolution is significantly associated with the instantaneous risk of death. To do that, we perform a Wald test on parameter α_{1k} and we note p the corresponding p -value. If $p \leq 0.05$, the biomarker k is selected for the next; otherwise, the biomarker is excluded from the analysis.

A sensibility analysis on the thresholds was conducted. Two alternative scenarios were considered: a stringent scenario and a flexible scenario. In the stringent scenario, the thresholds for the measurement errors and the RSE of parameters were reduced by 40% of their values. In the flexible scenario, they were increased by 40% of their values. The different cases are summarized in Table S3a.

3.2 | Multivariate modeling

The biomarkers selected at the univariate stage are grouped following the initial classification (given in Table S1). For each group, the biomarker with the most significant p -value for α_{1k} , and where there is at least one measurement in half of the patients, is considered for inclusion in the multivariate model. Thus, at most eight biomarkers are selected for the first step of the multivariate model that contains all these biomarkers. Then a first backward selection process including the selected biomarkers is performed. At each iteration, we remove the biomarker with the highest Wald test p -value for α_{1k} . We stop when all p -values for the coefficients α_{1k} are under 5%. If exists a biomarker for which the standard error of α_{1k} has not been estimated for convergence reasons, it is also removed. Finally, a second backward selection is performed by removing the biomarker with the highest Wald test p -value for α_{2k} and stop, as previously, when all p -values for the coefficients α_{2k} are significant. At the end of this step, K' biomarkers whose evolution is significantly associated with both instantaneous risk of death and discharge are selected.

For a given iteration where K'' biomarkers are involved in the modeling, the multivariate joint model writes:

$$\begin{aligned}
 y_{ij1} &= m_1(t_{ij1}, \psi_{i1}) + g[m_1(t_{ij1}, \psi_{i1}), \sigma_1] \epsilon_{ij} \\
 &\dots \\
 y_{ijK''} &= m_{K''}(t_{ijK''}, \psi_{iK''}) + g[m_{K''}(t_{ijK''}, \psi_{iK''}), \sigma_{K''}] \epsilon_{ij} \\
 \lambda_{1i}(t|\psi_i, \text{Score}_i) &= h_1 \exp \left[\sum_{k=1}^{K''} (\alpha_{1k} \times (m_k(t, \psi_{ik}) - \text{med}_k)) + \beta_1 \cdot \text{Score}_i \right] \\
 \lambda_{2i}(t|\psi_i, \text{Score}_i) &= h_2 \exp \left[\sum_{k=1}^{K''} (\alpha_{2k} \times (m_k(t, \psi_{ik}) - \text{med}_k)) + \beta_2 \cdot \text{Score}_i \right]
 \end{aligned} \tag{6}$$

with $\psi_i = (\psi_{i1}^T, \dots, \psi_{iK''}^T)^T$. As for the univariate model, $\lambda_{1i}, \lambda_{2i}$ are the subdistribution hazards for the risk of in-hospital death and discharge from hospital, h_1, h_2 the constant baseline risks, and β_1, β_2 the baseline score effects.

Here, the likelihood writes:

$$L(\theta) = \prod_{i=1}^N \int_{\eta_i} \left[\prod_{k=1}^{K''} \prod_{j=0}^{n_{ik}} p(y_{ijk} | \eta_{ik}, \text{Score}_i; \theta) \right] p(\bar{T}_i, \delta_i | \eta_i; \theta) p(\eta_i; \theta) d\eta_i, \tag{7}$$

where $\theta = \{\mu, \Omega, \sigma, h_1, h_2, \alpha_1, \alpha_2, \beta_1, \beta_2\}$ is the vector of parameters to estimate, with $\mu = (\mu_1^T, \dots, \mu_{K''}^T)^T$, $\Omega = \text{diag}(\omega_1^2, \dots, \omega_{K''}^2)$, $\sigma = (\sigma_1^T, \dots, \sigma_{K''}^T)^T$, $\alpha_1 = (\alpha_{11}, \dots, \alpha_{1K''})^T$, $\alpha_2 = (\alpha_{21}, \dots, \alpha_{2K''})^T$.

Same definitions of $p(\eta_i; \theta)$ and $p(t, \delta | \eta_i; \theta)$ as presented in Section 3.1.1 are used. Similar to the univariate joint models, the multivariate joint models are estimated by the SAEM algorithm on Monolix software, with the same settings as previously described.

4 | DYNAMIC PREDICTIONS AND PERFORMANCES COMPARISON

4.1 | Dynamic predictions

For patient i who has longitudinal observations for biomarker k until a landmark time l : $Y_{ik}(l) = \{y_{ijk}; 0 \leq t_{ij} \leq l\}$, we aim at predicting its future biomarker k values $m_{ik}(l + t|l)$ and his associated survival probability $s_i(l + t|l)$ where t is the horizon time. Since the dynamic predictions are applied for patients who are still hospitalized at time l , we focus on the following conditional probability:

$$s_i(l + t|l) = \mathbb{P}(T_{1i} > l + t | T_i > l, Y_{ik}(l), \text{Score}_i, \theta).$$

For each landmark time l , the biomarker measurements of patient i up to time l and the information provided by the population parameters estimated on the whole data are used to compute the a posteriori distribution of the individual parameters and infer the desired quantities. Monte-Carlo process as described in literature (Lavalley-Morelle et al., 2022) is adapted to derive 200 samples of individual parameters. Briefly, the probability of not experiencing the death between l and $l + t$, for a replicate r and an individual i , is computed as follows:

$$s_i^{(r)}(l + t | \psi_i^{(r)}) = \frac{F_{1i}^{(r)}(\infty) + F_{2i}^{(r)}(\infty) - F_{1i}^{(r)}(l + t) - F_{2i}^{(r)}(l)}{F_{1i}^{(r)}(\infty) + F_{2i}^{(r)}(\infty) - F_{1i}^{(r)}(l) - F_{2i}^{(r)}(l)} \tag{8}$$

and estimates of $m_{ik}(l + t|l)$, $k = 1, \dots, K'$, and $s_i(l + t|l)$ are obtained by taking the median over 200 replicates. Prediction intervals are obtained by reporting the 2.5th and the 97.5th percentiles of the distribution.

TABLE 2 Parameter estimates for the baseline model.

Parameter (unit)	Value	SE	RSE (%)	p-Value
Death				
$h_1(d^{-1})$	0.0003	0.00016	53	
β_1	0.36	0.05	14	$<10^{-5}$
Discharge				
$h_2(d^{-1})$	0.13	0.016	13	
β_2	-0.14	0.02	13	$<10^{-5}$

Note: p-Value of Wald tests are provided for the coefficients associated to the baseline score effect.

Abbreviations: RSE, relative standard error; SE, standard error.

4.2 | Comparison with the baseline model

For the baseline model, two subdistribution hazard functions are considered (for the risk of in-hospital death and discharge from hospital). Both are adjusted on the value of the baseline score:

$$\lambda_{1i}(t) = h_1 \exp(\beta_1 \cdot \text{Score}_i) \quad (9)$$

$$\lambda_{2i}(t) = h_2 \exp(\beta_2 \cdot \text{Score}_i).$$

For both baseline and multivariate joint modeling, three landmark times are considered at 3, 6, and 9 days after patient admission. Multiple horizon times are also considered, from landmark time plus 1–30 days after admission. The predictive performances are assessed and compared by computing the time-dependent receiver operator characteristic area under the curve (ROC AUC) for both the multivariate joint model and the baseline model. Details on computation are given in Supporting Information S1. We consider the cumulative–dynamic definition described in previous works (Blanche et al., 2015) using the *timeROC* (Blanche et al., 2013) package available in R software. Finally, tests for comparing the time-dependent ROC AUC of both models is also computed using the *timeROC* package.

5 | RESULTS

5.1 | Baseline model

Table 2 presents parameter estimates for the baseline model. A significant effect of the baseline score on both risks is highlighted: a patient with a one point higher baseline score has his risk of death multiplied by 1.43 (95% CI [1.30, 1.58]) and his risk of discharge divided by 1.15 (95% CI [1.11, 1.20]).

5.2 | Uni- and multivariate joint models

Figure 2 details the selection process of the 59 biomarkers. Based on the value of the residual error of the longitudinal submodels, 31 biomarkers are considered with a linear joint model and 22 with a nonlinear joint model. For the 31 linear biomarkers, 2 are excluded because at least one RSE of a survival parameter is over 100%, and 7 excluded because the link between the longitudinal and the survival part is not significant. For the nonlinear biomarkers, 19 are excluded because at least one RSE of a longitudinal or a survival parameter is over 100%. Hence, 22 biomarkers are selected with a linear model and 3 with a nonlinear model.

Table S4 presents those 25 biomarkers grouped following the initial classification and ranked by increasing p-value Wald test on α_{1k} . Seven biomarkers are selected for the multivariate analysis: blood neutrophil counts, D-dimers, arterial pH, lactate dehydrogenase (LDH), albuminemia, CRP (in logarithmic scale), and NT-proBNP.

After backward selection, we selected three biomarkers significantly associated with both risks of in-hospital death and discharge: the blood neutrophil counts, the arterial pH, and the CRP. Figure 3 shows the longitudinal evolution of

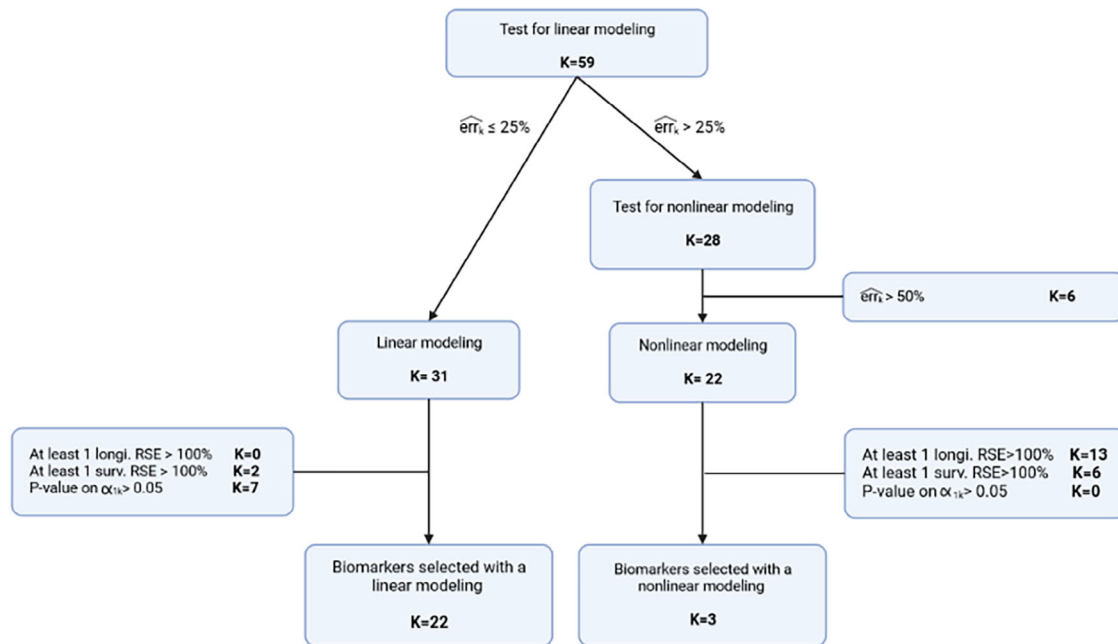


FIGURE 2 Flowchart for the univariate selection.

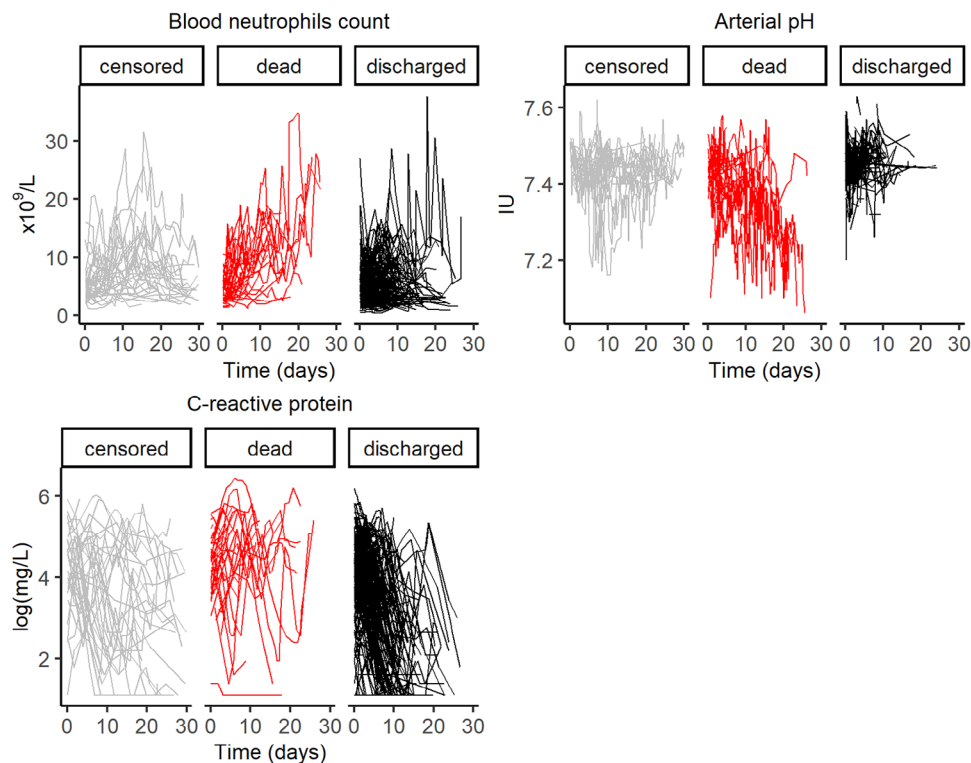


FIGURE 3 Individual longitudinal evolution of the selected biomarkers, stratified by clinical outcome.

each biomarker. Parameter estimates for the multivariate joint model are shown in Table 3. An increase in the blood neutrophil counts or the CRP, or a decrease in arterial pH is associated with a higher risk of death ($\alpha_{1n} = 0.14$, 95% CI [0.07, 0.20], $\alpha_{1c} = 0.63$, 95% CI [0.20, 1.06], $\alpha_{1p} = -11.4$, 95% CI [-16.93, -5.87]) and a lower risk of discharge ($\alpha_{2n} = -0.14$, 95% CI [-0.25, -0.03], $\alpha_{2c} = -1.09$, 95% CI [-1.44, -0.74], $\alpha_{2p} = 25.2$, 95% CI [12.75, 37.65]). Even after adjustment on the biomarker levels, the baseline score also has a significant effect on both risks: a higher baseline score is associated with

TABLE 3 Parameter estimates of the multivariate joint model.

Parameter (unit)	Value	SE	RSE (%)	p-Value
Longitudinal submodel				
Blood neutrophil counts				
μ_{0n} ($10^9 \cdot L^{-1}$)	4.59	0.16	3.6	
μ_{1n} ($10^9 \cdot L^{-1} \cdot d^{-1}$)	-0.15	0.024	16.0	
μ_{2n} ($10^9 \cdot L^{-1} \cdot d^{-1}$)	-0.16	0.016	10.0	
μ_{an} ($10^9 \cdot L^{-1}$)	5.30	1.58	29.8	
ω_{0n} ($10^9 \cdot L^{-1}$)	2.10	0.15	7.1	
ω_{1n} ($10^9 \cdot L^{-1} \cdot d^{-1}$)	0.13	0.017	13.0	
ω_{2n} ($10^9 \cdot L^{-1} \cdot d^{-1}$)	0.076	0.018	24.2	
ω_{an} ($10^9 \cdot L^{-1}$)	0.83	0.21	24.8	
σ_{bn}	0.32	0.0070	2.2	
Arterial pH				
μ_{0p}	7.44	0.0035	0.05	
μ_{1p} (d^{-1})	0.0027	0.00079	29.1	
ω_{0p}	0.039	0.0027	7.1	
ω_{1p} (d^{-1})	0.0053	0.00061	11.6	
σ_{ap}	0.055	0.00092	1.7	
C-reactive protein				
μ_{0c} ($\log(\text{mg} \cdot L^{-1})$)	4.18	0.061	1.5	
μ_{1c} ($\log(\text{mg} \cdot L^{-1}) \cdot d^{-1}$)	-0.16	0.012	7.3	
ω_{0c} ($\log(\text{mg} \cdot L^{-1})$)	0.93	0.048	5.2	
ω_{1c} ($\log(\text{mg} \cdot L^{-1}) \cdot d^{-1}$)	0.15	0.011	7.4	
σ_{ac} ($\log(\text{mg} \cdot L^{-1})$)	0.71	0.015	2.1	
Survival submodel				
Death				
h_1 (d^{-1})	0.00037	0.00024	65.3	
α_{1n} ($L \cdot 10^{-9}$)	0.14	0.033	24.2	$< 10^{-5}$
α_{1p}	-11.4	2.82	24.9	$< 10^{-5}$
α_{1c} ($-\log(\text{mg} \cdot L^{-1})$)	0.63	0.22	34.7	0.004
β_1	0.33	0.060	18.1	$< 10^{-5}$
Discharge				
h_2 (d^{-1})	0.014	0.0069	51.4	
α_{2n} ($L \cdot 10^{-9}$)	-0.14	0.054	39.1	0.01
α_{2p}	25.2	6.35	25.2	$< 10^{-5}$
α_{2c} ($-\log(\text{mg} \cdot L^{-1})$)	-1.09	0.18	16.2	$< 10^{-5}$
β_2	-0.12	0.032	27.3	0.0002

Note: p-Values of Wald tests are reported for the link coefficients and the coefficients associated with the baseline score effect.

Abbreviations: RSE, relative standard error; SE, standard error.

a higher risk of death ($\beta_1 = 0.33$, 95% CI [0.21, 0.45]) and a lower risk of discharge ($\beta_2 = -0.12$, 95% CI [-0.18, -0.05]). Some diagnostic plots that assessed the quality of the longitudinal submodels can be found in Figure S1.

The results of uni- and multivariate selection under stringent and flexible scenarios of the sensibility analysis are described in Table S3b and Figures S2a and S2b. When stringent thresholds are used, the interest of the backward strategy is limited because only three biomarkers entered in the multivariate analysis and remained in the final model. Of note, those three biomarkers are not selected in the current scenario. On the contrary when thresholds are too flexible, more biomarkers are involved and the final multivariate joint model retains two biomarkers that are also selected in the cur-

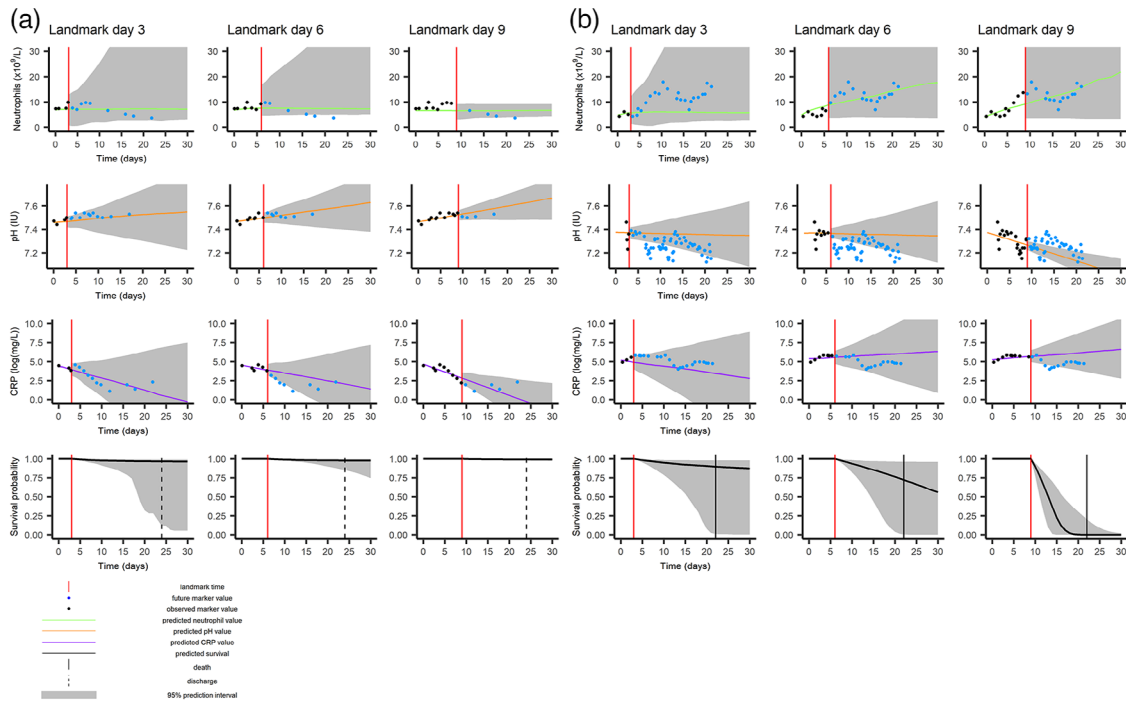


FIGURE 4 Dynamic predictions and 95% prediction intervals for patient A (left) and patient B (right). Patient A was discharged at time 24 and patient B died at time 22. Predictions for biomarker evolution are on the first three rows (blood neutrophils count, arterial pH, and C-reactive protein (CRP), respectively), and predictions for the survival on the last row.

rent scenario (pH and CRP). Of note, under the flexible scenario, blood neutrophil counts are modeled linearly and are removed during the backward process.

Finally, some details about computation time are given in Supporting Information S2. We evaluated through simulations the CPU times needed to estimate parameters as well as standard errors. Estimation of both parameters and standard errors is scalable according to the number of patients included (the estimation process will take twice as long if there are twice as many subjects). However, the relation between the CPU time spent to estimate standard errors and the number of biomarkers involved is almost exponential. Considering a too large set of biomarkers in the backward process would be inappropriate with this approach.

5.3 | Dynamic predictions and performances comparison

First, to illustrate the dynamic predictions, we consider two patients. Patient A, who has a baseline score of 8, was discharged at day 24. Patient B, who has a baseline score of 6, died at day 22. We first computed survival probabilities from the baseline model for these two patients at horizon time 30. We predicted a high survival probability for both patients: 90% (95% CI [87, 93]) and 95% (95% CI [93, 97]) for patients A and B, respectively. Of note, patient B has a greater probability of discharge because of a lower baseline score. Figure 4 shows the dynamic longitudinal and survival predictions from the full multivariate joint model for the two patients, for different landmark and horizon times. For patient A, all the predictions are good, in light with the result of the baseline model and the actual outcome for this patient. For patient B, we see that his survival probability is getting worse as the landmark time increases. He finally died at time 22. In this case, the use of the full information available up to the landmark time helps to improve predictions.

To assess the prediction performances of the model, time-dependent ROC AUC are computed for the baseline model as well as the multivariate joint model. Figure 5 shows the results of this estimation. Given the small sample of failures for the times close to the landmark time, we are interested in predictions for late horizon times. The multivariate joint model has higher ROC AUC for late horizon times, compared to the baseline model. For example, for landmark time 9 and horizon time 30, AUC [95% CI] is 0.64 [0.55, 0.74], and 0.84 [0.75, 0.93] for the baseline and the joint model, respectively (Table 4). There is a significant difference in ROC AUC between both models for landmark times 6 and 9. In other words,

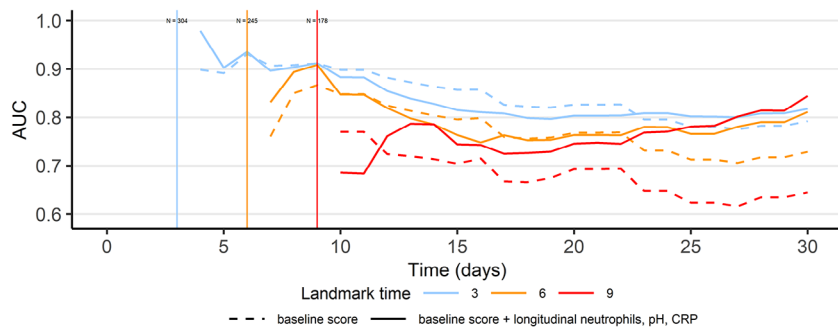


FIGURE 5 Time-dependent area under the curve (AUC) computed on the data set for various landmark and horizon times. The number of at-risk patients at each landmark is given on the top.

TABLE 4 Tests for comparing ROC AUC at horizon time 30 between the baseline and the joint model.

Horizon time = 30	Landmark day 3	Landmark day 6	Landmark day 9
AUC [95% CI]—baseline model	0.79 [0.65,0.81]	0.73 [0.65,0.81]	0.64 [0.55, 0.74]
AUC [95% CI]—joint model	0.82 [0.75,0.89]	0.81 [0.73,0.89]	0.84 [0.75,0.93]
p-value	0.37	0.04	$< 10^{-5}$

Abbreviation: CI, confidence interval; ROC AUC, receiver operator characteristic area under the curve.

when sufficient information is available, the multivariate joint model performs better compared to the baseline model and the benefit is higher as more information becomes available.

Of note, we also computed ROC AUC for multivariable joint model under stringent and flexible scenarios of the sensitivity analysis. ROC AUC [95% CI] for landmark = D9 and horizon = D30 are, respectively, 0.76 [0.67, 0.85] and 0.81 [0.72, 0.90] for the stringent and the flexible scenario. This reduction, compared with our current scenario, is not significant (p -values are, respectively, 0.05 and 0.09). Details are given in Table S3b.

Finally, to show the interest of biomarker selection, we present the results of the multivariate joint model estimation including the seven biomarkers considered in the multivariate analysis in Table S5. Some biomarker effects disappear after adjustments on all others and the complexity of estimation reflects a high degree of uncertainty particularly in the estimates of link coefficients. The ROC AUC is not significantly improved with the seven biomarkers (ROC AUC [95% CI] for landmark = D9 and horizon = D30 is 0.85 [0.77, 0.93], $p = 0.50$).

6 | SIMULATION STUDY

We performed a simulation study to answer three points. The first is to show that the SAEM algorithm implemented on Monolix software provides good estimation performances of such multivariate joint model. The second is to assess the performances of the selection process at the multivariate stage (i.e., the backward process). Finally, we will assess the ROC AUC indicator under different scenarios because some authors described a possible overestimation of this indicator (Schmid et al., 2013).

6.1 | Simulation settings

We simulated $M = 100$ data sets of $N = 300$ patients assuming that $K = 7$ biomarkers (bm_1 to bm_7) were available at most 30 days after the hospital admission. The biomarkers were simulated according to the design of the application, that is to say considering the patterns and the estimated parameters of the seven biomarkers included in the multivariate analysis. Biomarker measurements were reported for each patient at different frequencies until time to in-hospital death or discharge from hospital or also end of the first 30 days of the stay, with no other censoring process. Table 5 summarizes those characteristics. Longitudinal submodels refer to Equations (2) and (3) with $t_{lag} = 0$.

For the survival part, we considered two competing events. We supposed that bm_1 , bm_2 , and bm_3 are truly associated with the event 1 (risk of in-hospital death), bm_4 and bm_5 are not associated but have their slopes correlated with those of bm_2 and bm_3 , respectively, and bm_6 and bm_7 are not associated and not correlated with other biomarkers. A scenario with stronger correlations was also considered. In this one, we simulated bm_4 and bm_5 having a correlation on intercept

TABLE 5 Characteristics of the seven simulated biomarkers.

Biomarker	Correspondence	Longitudinal submodel	Error model	Measurement frequency
bm_1	Blood neutrophil counts	Nonlinear	Proportional	Every 2 days
bm_2	Arterial pH	Linear	Additive	Every 1.5 days
bm_3	CRP	Linear	Additive	Every 2 days
bm_4	NT-proBNB	Linear	Additive	Every 3 days
bm_5	D-dimers	Linear	Proportional	Every 3 days
bm_6	Albuminemia	Linear	Proportional	Every 3 days
bm_7	LDH	Linear	Proportional	Every 3 days

Abbreviations: CRP, C-reactive protein; LDH, lactate dehydrogenase.

and slope parameters of 0.95 with the ones of bm_2 and bm_3 , respectively. We followed methods already described (Fine, 2001; Lavalley-Morelle et al., 2022) to generate failure time data. The subdistribution for the risk of in-hospital death is given by

$$\mathbb{P}(T_{1i} < t | \psi_i, W_i; \theta) = 1 - \exp\left(-\int_0^t \frac{p_1 g_1 \exp(-g_1 \times s) \times \lambda(s)}{1 - p_1(1 - \exp(-g_1 \times s))} ds\right) \quad (10)$$

with $\lambda(s) = \exp[\alpha_{11} \times (m_1(\psi_{i1}, t_{ij1}) - med_1) + \alpha_{12} \times (m_2(\psi_{i2}, t_{ij2}) - med_2) + \alpha_{13} \times (m_3(\psi_{i3}, t_{ij3}) - med_3)]$. Parameters p_1 and g_1 are used to control the number of event 1 at infinite time and the speed of onset.

The subdistribution for the competing risk was then obtained using an exponential distribution with rate $\frac{t}{b}$ where b controls the speed of event 2 onset:

$$\mathbb{P}(T_{2i} < t | \psi_i, W_i; \theta) = (1 - F_1(\infty)) \times \left(1 - \exp\left(-\frac{t}{b}\right)\right). \quad (11)$$

p_1, g_1 were set to 0.05, 0.1, respectively, and b was fixed to 10 to obtain about 17% cause 1 failures, 69% cause 2 failures, and 14% administrative censoring at time 30, similar to the application. Further details about true values for model parameters are given in Supporting Information S3.

6.2 | Evaluation of the estimation performances, selection process, and ROC AUC indicator

6.2.1 | Evaluation of the estimation performances

For each simulated data set, the true model is estimated and writes as follows, considering $K'' = 3$:

$$\begin{aligned} y_{ij1} &= m_1(t_{ij1}, \psi_{i1}) + g[m_1(t_{ij1}, \psi_{i1}), \sigma_1] \epsilon_{ij} \\ &\dots \\ y_{ijK''} &= m_{K''}(t_{ijK''}, \psi_{iK''}) + g[m_{K''}(t_{ijK''}, \psi_{iK''}), \sigma_{K''}] \epsilon_{ij} \\ \lambda_{1i}(t | \psi_i) &= \frac{p_1 g_1 \exp(-g_1 \times t)}{1 - p_1(1 - \exp(-g_1 \times t))} \exp\left[\sum_{k=1}^{K''} (\alpha_{1k} \times (m_k(t, \psi_{ik}) - med_k))\right] \\ \lambda_{2i}(t | \psi_i) &= \frac{p_2 g_2 \exp(-g_2 \times t)}{1 - p_2(1 - \exp(-g_2 \times t))} \exp\left[\sum_{k=1}^{K''} (\alpha_{2k} \times (m_k(t, \psi_{ik}) - med_k))\right]. \end{aligned} \quad (12)$$

The model is estimated by maximization of the likelihood using the SAEM algorithm implemented on Monolix 2018 software with the same settings as explained in the application.

We aim to assess the performances of the estimation on the longitudinal parameters and the survival parameters for the event 1. We define $\theta = \{\mu, \omega, p_1, g_1, \alpha_{11}, \alpha_{12}, \alpha_{13}\}$ and $\hat{\theta}$ its estimate. The model specification for data simulation does not allow to easily identify $p_2, g_2, \alpha_{21}, \alpha_{22},$ and α_{23} that is why estimation evaluation was based on vector $\hat{\theta}$. We evaluated the empirical relative bias and relative root mean square errors (RRMSE) of $\hat{\theta}$ defined as:

$$\text{Relativebias} = \frac{1}{M} \sum_{m=1}^M \frac{\hat{\theta}^m - \theta}{\theta} \times 100 \quad \text{and} \quad \text{RRMSE} = \sqrt{\frac{1}{M} \sum_{m=1}^M \left(\frac{\hat{\theta}^m - \theta}{\theta} \times 100 \right)^2}. \quad (13)$$

We also provided violin plots of relative estimation errors (REE) defined for each data set $m \in \{1, \dots, M\}$ as:

$$\text{REE}^m = \frac{\hat{\theta}^m - \theta}{\theta} \times 100. \quad (14)$$

6.2.2 | Evaluation of the backward process

The objective is to assess the ability of the backward process to find the “true” model, that is to say the final set of biomarkers truly associated with the risk of in-hospital death. The first iteration of the process starts with the seven biomarkers included in the multivariate joint model, and at each iteration, the biomarker with the highest p -value on α_{1k} is removed. We stop when all the p -values for $\alpha_{1k} < 5\%$. For each data set $m \in \{1, \dots, M\}$ and for a given iteration of the backward process involving K'' biomarkers, we estimated the joint model defined in Equation (12).

We finally report the final set of biomarkers after the backward process for each simulation and each scenario of correlations.

6.2.3 | Evaluation of the ROC AUC indicator

To answer this objective, we computed the ROC AUC for the landmark time D9 and horizon time D30 for both true model and final model provided at the end of the backward process, for each simulated data set. We compared both distributions using paired t -test.

6.3 | Results

6.3.1 | Evaluation of the estimation performances

Table 6 provides relative bias and RRMSE expressed in percentage. Relative bias and RRMSE are low for most of parameters. Longitudinal parameters are accurately estimated (while more uncertainty for some random effects, with higher relative bias and RRMSE). The parameters that link the longitudinal and the survival processes ($\alpha_{11}, \alpha_{12}, \alpha_{13}$) are of particular interest and are estimated with a relative bias around 13% and an RRMSE around 30%. The design of the simulation may cause this moderate uncertainty of estimation for those parameters. In a previous work where only one biomarker was handled in a richer design and with a higher number of patients (Lavalley-Morelle et al., 2022), the RRMSE were reduced. Overall, this suggests good estimation performances given the size of the data set and in particular, the small number of event 1 (deaths). Violin plots can be found in Figure S3 and provide same conclusion.

6.3.2 | Evaluation of the backward process

The upset plot on Figure 6 is an efficient way to visualize the proportion of replicates finding the truly associated biomarkers. In the first scenario of correlations (correlation of 0.8 on slope parameters), 75% of the replicates contains the three correct biomarkers ($bm_1, bm_2,$ and bm_3): 64% only those three, 9% with one additional biomarker, and 2% with two additional biomarkers. In 21% of cases, the process found only two of the true associated biomarkers. For the remaining cases,

TABLE 6 Estimated relative bias and root mean square errors (RMSE) based on 100 simulations.

Parameter	True value	Relative bias (%)	RRMSE (%)
Longitudinal part—biomarker 1			
Fixed effects			
μ_{01}	4.6	0.4	3
μ_{11}	-0.15	-8	15
μ_{21}	-0.16	-7	14
μ_{a1}	5.3	-9	17
Random effects			
ω_{01}	2.0	-2	11
ω_{11}	0.10	14	28
ω_{21}	0.07	-8	36
ω_{a1}	0.80	25	46
Error parameter			
σ_{b1}	0.30	2	3
Longitudinal part—biomarker 2			
Fixed effects			
μ_{02}	7.4	3^{-5}	0.04
μ_{12}	0.003	-6	18
Random effects			
ω_{02}	0.04	2	11
ω_{12}	0.005	6	17
Error parameter			
σ_{a2}	0.05	-0.1	1
Longitudinal part—biomarker 3			
Fixed effects			
μ_{03}	4.2	-0.08	1
μ_{13}	-0.16	-5	9
Random effects			
ω_{03}	0.90	-0.9	11
ω_{13}	0.15	7	15
Error parameter			
σ_{a3}	0.70	-0.1	2
Survival part			
p_1	0.05	-23	37
g_1	0.10	16	32
α_{11}	0.14	15	35
α_{12}	-11	13	32
α_{13}	0.60	11	23

Abbreviation: RRMSE, relative root mean square errors.

the process only found one associated biomarker (2%) or bm_4 (correlated with bm_2) was selected instead of bm_2 (2%). We emphasize that all final models should have similar predictive performances.

For the scenario of stronger correlations, the upset plot is given in Figure S4. Similar to the previous scenario of correlations, 72% of the replicates contains a correct set of biomarkers: 43% the true set of biomarkers, 15% with a very correlated biomarker instead of the associated one, 13% with one additional biomarker, and 1% with three additional biomarkers. In 20% of cases the backward selected only two associated biomarkers, and in 4%, only one. For the remaining cases (4%), the process retains one or two associated biomarkers with an additional one.

Overall, the backward strategy demonstrates good performances even in the presence of strong correlations.

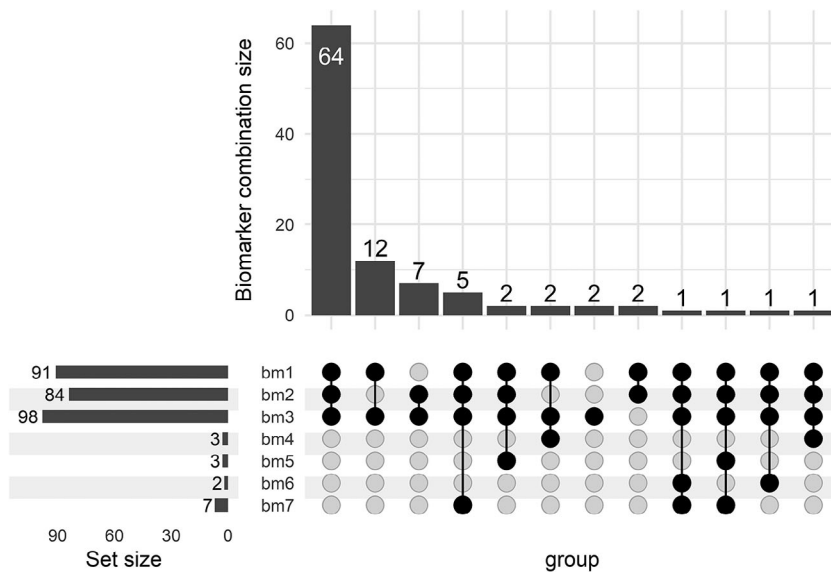


FIGURE 6 Final set of biomarkers selected after the backward process (simulation study).

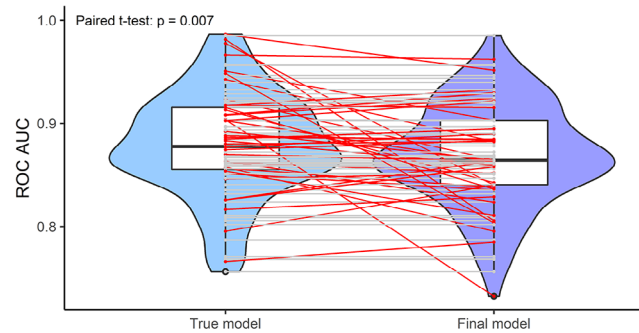


FIGURE 7 Distributions of receiver operator characteristic area under the curve (ROC AUC) of the true model and the final model for landmark time = D9 and horizon time = D30. True model corresponds to the multivariate joint model with bm1, bm2, and bm3 and the final model the one with the biomarkers selected after the backward process. Gray lines represent situations when the final model is equal to the true model and red lines represent other cases.

6.3.3 | Evaluation of the ROC AUC indicator

Figure 7 provides the distribution of ROC AUC for landmark = D9 and horizon time = D30 under the true and the final model, for the first scenario of correlations. Of note, 64% of ROC AUC were equal for both models, as the backward strategy leads to the true set of biomarkers in those cases. The figure shows that ROC AUC, in most cases, tends to be lower when a wrong model is selected, and more rarely, overestimated. In mean [95% CI], the ROC AUC is reduced by 0.010 [0.003, 0.017] ($p = 0.007$) under the final model, which, although significant, is very small.

7 | DISCUSSION

In this work, we proposed to evaluate the added value of modeling the full trend of various biomarkers to predict the risk of in-hospital death for COVID-19 patients, compared to using only information available at hospital admission. We developed a strategy to select between various biomarkers most associated with prognostic. The proposed methodology has the advantage to select a subset of most relevant biomarkers based on statistical methodology and not only on clinical considerations. We explored more information than an entire clinical-based approach that would directly considered a small set of biomarkers in the final model. However, we emphasize that building a predictive tool according to statistical criteria or only clinical knowledge may produce similar results.

We found that the dynamics of the blood neutrophil counts, arterial pH and CRP were significantly associated with the risk of in-hospital death and discharge from hospital: an increase in the blood neutrophil counts or the CRP, or a decrease in the pH is associated with a higher risk of death and a lower risk of discharge. We showed that the individual predictions from the multivariate joint model were significantly improved and outperformed those from the baseline model when sufficient information was available. Hence, such a tool can be useful for clinicians in routine medical practice as it may support clinical decisions such as therapeutic escalation or limitation of care. These decisions usually occur during the follow-up, and considering up-to-date longitudinal information to derive survival probabilities seems crucial. More generally, our methodology could be straightforwardly extended to other outcomes or medical conditions, and be easily implemented into hospital information systems to provide clinicians with dynamically updated prognostic scores and their estimated uncertainty.

Our real-case study also has some limitations. First, we recall that the baseline score referring to the 4C score was computed for each patient with some missing data for CRP and urea components. While it affects few individuals, the single imputation of missing data, given the other variables of the score, can introduce a small bias in the modeling, damped by the fact that the entire score slightly depends only slightly on those components. Second, one can argue that complex biomarker trends are not accounted for in the choice of the modeling. However, in such case, we think that no parametric model can approximate the evolution, and nonparametric models are out of the scope of this article. Moreover, it is worth noting that biomarkers whose evolution is hard to model are generally noninformative biomarkers that have no added value in terms of predictions. Then, two different thresholds for the residual error were used for the selection of linear and nonlinear models during univariate analysis. This relies on subjective appreciations and other thresholds could be proposed as well. Some biological measurements exhibit erratic behavior and are hard to model: The residual error should not be too high, but some flexibility is needed. We showed in the sensibility analysis that the choice of the thresholds in the univariate analysis impacts the set of biomarkers selected in the final model. However, all the scenarios considered lead to predictive performances that do not differ significantly. We also considered a set of biomarkers frequently prescribed to avoid at most selective missingness. However, it is worth noting that some biomarkers may have been more often measured in severe patients. In an extreme case, the population parameters estimated for such biomarker can be biased and not reflecting the dynamic in the whole population. In the survival part, the baseline risk functions considered may seem restrictive. However, the objective of the work is to provide a powerful predictive tool and we showed through the application that the predictions performed well with this parametrization. That is why no more complex function was considered. Then, in the backward processes, we chose to select biomarkers that are associated with both competing risks. As an alternative approach and for other applications, we could allow for different biomarkers being associated with one or both risks. Finally, our model should also be validated on an external data set before being used in practice. Due to the sample size and the number of events, an internal validation by splitting data was found inappropriate.

In a methodological point of view, we demonstrated well estimation performances of the SAEM algorithm. We evaluated the estimation in a complex design with three longitudinal submodels (with one nonlinear) and two competing risks with a small number of event 1. Bias and RRMSE were relatively low for most of parameters. However, as we are in a subdistribution approach, estimation performances relative to survival parameters associated with event 2 are impossible to evaluate. It may be an interesting future research. Moreover, although backward selection procedure has known drawbacks, it has never been proposed in the context of joint modeling with competing risks and we found here interesting performances in the simulation study even when introducing correlations between biomarkers. We emphasize that strong correlation structures were considered in the simulations. In reality, correlations strongly depend on measurement errors and are probably lower than those considered. Other selection strategies as the forward or forward-backward stepwise strategies may be considered, although they may be more time-consuming because of the number of combinations to test at each iteration of the process. We acknowledge that the backward strategy may be computationally expensive depending on the machine hardware and software configuration. We showed through simulations that this strategy may be extended for a larger set of patients but not for a too large set of biomarkers. In fact, the stochastic algorithm used to compute the FIM shows an exponential growth in terms of CPU time spent as the number of biomarkers increases. This algorithm is implemented according to the Louis method and is run once parameters have been estimated (see Kuhn and Lavielle, 2005, for more details). The whole estimation process (parameters and standard errors) is thus, time-consuming. Recently, Delattre and Kuhn (2023) developed an interesting alternative: a stochastic algorithm to compute the FIM as the SAEM algorithm iterates. It offers the potential to be less time-consuming, however, it has not been evaluated for multiple responses and joint models so future developments are needed to consider this alternative. Finally, we note that the *compare* function used for testing and comparing time-dependent AUC between the baseline and the multivariate joint model was not designed for nested models. It would imply a loss of power and not a rise of the type-I error. Hence, there is no a negative impact on our results.

A possible extension of this work could be the implementation of a selection procedure based on penalized likelihood such as LASSO penalization because forward/backward selection on regression models have been shown to be outperformed by this type of methods in a high-dimensional context.

ACKNOWLEDGMENTS

The authors thank Dr. Arthur Mageau, AP-HP, Bichat Claude Bernard University Hospital, F-75018 Paris, France, for his clinical expertise and help.


CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Research data are not shared. We provide the code to reproduce each result of the article (with a jittered version of the true data set for the application part) at <https://github.com/alexandralavmo/BiometricalJournal/tree/main/codenoIR>. As the computation tasks may be very time-consuming, we also provide a version including intermediate results available at <https://github.com/alexandralavmo/BiometricalJournal/tree/main/codeIR>. The commit version number: 321cdb6d3e0469fb25a6dfdd41bd57837f6de84c.

OPEN RESEARCH BADGES

 This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the [Supporting Information](#) section.

This article has earned an open data badge “**Reproducible Research**” for making publicly available the code necessary to reproduce the reported results. The results reported in this article were reproduced partially due to data confidentiality issues.

ORCID

Alexandra Lavalley-Morelle  <https://orcid.org/0000-0001-5485-5694>

Simon B. Gressens  <https://orcid.org/0000-0002-4993-6956>

REFERENCES

- Andrinopoulou, E.-R., Rizopoulos, D., Takkenberg, J. J., & Lesaffre, E. (2017). Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. *Statistical Methods in Medical Research*, 26(4), 1787–1801. <https://doi.org/10.1177/0962280215588340>
- Blanche, P., Dartigues, J., & Jacqmin-Gadda, H. (2013). Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in Medicine*, 32(30), 5381–5397. <https://doi.org/10.1002/sim.5958>
- Blanche, P., Proust-Lima, C., Loubère, L., Berr, C., Dartigues, J., & Jacqmin-Gadda, H. (2015). Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics*, 71(1), 102–113. <https://doi.org/10.1111/biom.12232>
- Delattre, M., & Kuhn, E. (2023). *Estimating Fisher information matrix in latent variable models based on the score function*. <https://hal.science/hal-02285712>
- Delyon, B., Lavielle, M., & Moulines, E. (1999). Convergence of a stochastic approximation version of EM algorithm. *Annals of Statistics*, 27, 94–128.
- Desmée, S., Mentré, F., Veyrat-Follet, C., Sébastien, B., & Guedj, J. (2017). Nonlinear joint models for individual dynamic prediction of risk of death using Hamiltonian Monte Carlo: Application to metastatic prostate cancer. *BMC Medical Research Methodology*, 17(1), 105. <https://doi.org/10.1186/s12874-017-0382-9>
- Fine, J. P. (2001). Regression modeling of competing crude failure probabilities. *Biostatistics (Oxford, England)*, 2(1), 85–97. <https://doi.org/10.1093/biostatistics/2.1.85>
- Jeong, J.-H., & Fine, J. (2006). Direct parametric inference for the cumulative incidence function. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 55(2), 187–200. <https://www.jstor.org/stable/3592662>
- Keroui, M., Mercier, F., Bertrand, J., Tardivon, C., Bruno, R., Guedj, J., & Desmée, S. (2020). Bayesian inference using Hamiltonian Monte-Carlo algorithm for nonlinear joint modeling in the context of cancer immunotherapy. *Statistics in Medicine*, 39(30), 4853–4868. <https://doi.org/10.1002/sim.8756>
- Knight, S. R., Ho, A., Pius, R., Buchan, I., Carson, G., Drake, T. M., Dunning, J., Fairfield, C. J., Gamble, C., Green, C. A., Gupta, R., Halpin, S., Hardwick, H. E., Holden, K. A., Horby, P. W., Jackson, C., Mclean, K. A., Merson, L., Nguyen-Van-Tam, J. S., ... Young, P. (2020). Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: Development and validation of the 4C mortality score. *British Medical Journal*, 370. <https://doi.org/10.1136/bmj.m3339>

- Kuhn, E., & Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, 49(4), 1020–1038. <https://doi.org/10.1016/j.csda.2004.07.002>
- Lavalley-Morelle, A., Timsit, J.-F., Mentré, F., Mullaert, J., & Network, T. O. (2022). Joint modeling under competing risks: Application to survival prediction in patients admitted in intensive care unit for sepsis with daily Sequential Organ Failure Assessment score assessments. *CPT: Pharmacometrics & Systems Pharmacology*, 11(11), 1472–1484. <https://onlinelibrary.wiley.com/doi/pdf/10.1002/psp4.12856>
- Lavillegrand, J.-R., Garnier, M., Spaeth, A., Mario, N., Hariri, G., Pilon, A., Berti, E., Fieux, F., Thietart, S., Urbina, T., Turpin, M., Darrivere, L., Fartoukh, M., Verdonk, F., Dumas, G., Tedgui, A., Guidet, B., Maury, E., Chantran, Y., ... Ait-Oufella, H. (2021). Elevated plasma IL-6 and CRP levels are associated with adverse clinical outcomes and death in critically ill SARS-CoV-2 patients: Inflammatory response of SARS-CoV-2 patients. *Annals of Intensive Care*, 11(9). <https://doi.org/10.1186/s13613-020-00798-x>
- Li, T., Wang, X., Zhuang, X., Wang, H., Li, A., Huang, L., Zhang, X., Xue, Y., Wei, F., & Ma, C. (2021). Baseline characteristics and changes of biomarkers in disease course predict prognosis of patients with COVID-19. *Internal and Emergency Medicine*, 16(5), 1165–1172. <https://doi.org/10.1007/s11739-020-02560-4>
- Liang, W., Liang, H., Ou, L., Chen, B., Chen, A., Li, C., Li, Y., Guan, W., Sang, L., Lu, J., Xu, Y., Chen, G., Guo, H., Guo, J., Chen, Z., Zhao, Y., Li, S., Zhang, N., Zhong, N., & He, J. (2020). Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Internal Medicine*, 180(8), 1–9. <https://doi.org/10.1001/jamainternmed.2020.2033>
- Liu, M., Sun, J., Herazo-Maya, J. D., Kaminski, N., & Zhao, H. (2019). Joint models for time-to-event data and longitudinal biomarkers of high dimension. *Statistics in Biosciences*, 11(3), 614–629. <https://doi.org/10.1007/s12561-019-09256-0>
- Long, J. D., & Mills, J. A. (2018). Joint modeling of multivariate longitudinal data and survival data in several observational studies of Huntington's disease. *BMC Medical Research Methodology*, 18(1), 138. <https://doi.org/10.1186/s12874-018-0592-9>
- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2), 226–233. <https://doi.org/10.1111/j.2517-6161.1982.tb01203.x>
- Mueller, A. A., Tamura, T., Crowley, C. P., DeGrado, J. R., Haider, H., Jezmir, J. L., Keras, G., Penn, E. H., Massaro, A. F., & Kim, E. Y. (2020). Inflammatory biomarker trends predict respiratory decline in COVID-19 patients. *Cell Reports. Medicine*, 1(8), 100144. <https://doi.org/10.1016/j.xcrm.2020.100144>
- Myrstad, M., Ihle-Hansen, H., Tveita, A. A., Andersen, E. L., Nygård, S., Tveit, A., & Berge, T. (2020). National Early Warning Score 2 (NEWS2) on admission predicts severe disease and in-hospital mortality from Covid-19—A prospective cohort study. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, 28(1), 66. <https://doi.org/10.1186/s13049-020-00764-3>
- Rajeswaran, J., Blackstone, E. H., & Barnard, J. (2018). Joint modeling of multivariate longitudinal data and competing risks using multiphase sub-models. *Statistics in Biosciences*, 10(3), 651–685. <https://doi.org/10.1007/s12561-018-9223-6>
- Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3), 819–829. <https://doi.org/10.1111/j.1541-0420.2010.01546.x>
- Schmid, M., Kestler, H. A., & Potapov, S. (2013). On the validity of time-dependent AUC estimators. *Briefings in Bioinformatics*, 16(1), 153–168. <https://doi.org/10.1093/bib/bbt059>
- Shen, F., & Li, L. (2021). Backward joint model and dynamic prediction of survival with multivariate longitudinal data. *Statistics in Medicine*, 40(20), 4395–4409. <https://doi.org/10.1002/sim.9037>
- Tjendra, Y., Al Mana, A. F., Espejo, A. P., Akgun, Y., Millan, N. C., Gomez-Fernandez, C., & Cray, C. (2020). Predicting disease severity and outcome in COVID-19 patients: A review of multiple biomarkers. *Archives of Pathology & Laboratory Medicine*, 144(12), 1465–1474. <https://doi.org/10.5858/arpa.2020-0471-SA>
- Tong-Minh, K., van der Does, Y., van Rosmalen, J., Ramakers, C., Gommers, D., van Gorp, E., Rizopoulos, D., & Endeman, H. (2022). Joint modeling of repeated measurements of different biomarkers predicts mortality in COVID-19 patients in the intensive care unit. *Biomarker Insights*, 17. <https://doi.org/10.1177/11772719221112370>
- Zhang, L., Yan, X., Fan, Q., Liu, H., Liu, X., Liu, Z., & Zhang, Z. (2020). D-dimer levels on admission to predict in-hospital mortality in patients with Covid-19. *Journal of Thrombosis and Haemostasis*, 18(6), 1324–1329. <https://doi.org/10.1111/jth.14859>
- Zhu, Z., Cai, T., Fan, L., Lou, K., Hua, X., Huang, Z., & Gao, G. (2020). The potential role of serum angiotensin-converting enzyme in coronavirus disease 2019. *BMC Infectious Diseases*, 20(1), 883. <https://doi.org/10.1186/s12879-020-05619-x>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lavalley-Morelle, A., Peiffer-Smadja, N., Gressens, S. B., Souhail, B., Lahens, A., Bounhiol, A., Lescure, F.-X., Mentré, F., & Mullaert, J. (2023). Multivariate joint model under competing risks to predict death of hospitalized patients for SARS-CoV-2 infection. *Biometrical Journal*, e2300049. <https://doi.org/10.1002/bimj.202300049>