Lights: a generalized joint model for high-dimensional multivariate longitudinal data and censored durations

Simon Bussy[*1], Antoine Barbieri[2], Sarah Zohar[1], and Anne-Sophie Jannot[1,3]

[1]*INSERM, UMRS 1138, Centre de Recherche des Cordeliers, Paris, France*
[2]*INSERM, UMR 1219, Bordeaux Population Health Research Center, Bordeaux, France*
[3]*Biomedical Informatics and Public Health Department, EGPH, APHP, Paris, France*

Abstract

This paper introduces a prognostic method called *lights* to deal with the problem of joint modeling of longitudinal data and censored durations, where a large number of longitudinal features are available. Yet there is no standard model so far to learn from such high-dimensional multivariate longitudinal data in a survival analysis setting. The statistical performance of the method is examined on an extensive Monte Carlo simulation study, and finally illustrated on a publicly available dataset. On these high-dimensional datasets, our proposed method significantly outperforms the state-of-the-art survival models regarding risk prediction in terms of C-index, with a computing time orders of magnitude faster. In addition, it provides powerful interpretability by automatically pinpointing significant features being relevant from a clinical perspective. Thus, we propose a powerfull tool for personalized medicine, with the ability of automatically determining significant prognostic longitudinal biomarkers, which is of increasing importance in many areas of medicine.

Keywords. High-dimensional estimation; Joint modeling; Multivariate longitudinal data; Survival analysis

1 Introduction

In many clinical studies, it has become increasingly common to record the values of longitudinal features (e.g. biomarkers) until the occurrence of an event of interest for a subject. The "joint modeling" approaches, namely modeling the longitudinal and survival outcomes through a joint likelihood model rather than separately, has received considerable attention during the past two decades [Wulfsohn and Tsiatis, 1997]. Numerical studies suggest that these approaches are among the most satisfactory to combine information [Tsiatis and Davidian, 2004, Yu et al., 2004]. They have the advantage of making more efficient use of the data since information about survival also goes into modeling the longitudinal features. In addition, they produce unbiased estimates and do not rely on approximations for incorporating conplex longitudinal trajectories.

The high-dimensional longitudinal data context. With the development of electronic health records, high-dimensional settings are becoming increasingly frequent in various contexts where the number of available features to consider as potential risk factors is tremendous. Moreover, with an increased focus on personalised medicine, the need to

---

[*]Corresponding author: simon.bussy@gmail.com

implement multivariate models that account for a large number of longitudinal outcomes is critical. Despite this, joint models have predominantly focused on univariate data, with attempts to fit multiple univariate joint models separately [Wang et al., 2012], which is inefficient [Lin et al., 2002]. Despite many multivariate models being presented in full generality, questions arrising from the high-dimensional context – computational power, limits in numerical estimation, or sample size for instance – are never considered in analyses (to the best of our knowledge), and the number of longitudinal outcomes considered in numerical studies are often very low (see Hickey et al. [2016] for a complete review). For instance, Jaffa et al. [2014] only considers 3 longitudinal outcomes in the simulation study while mentioning a "high-dimensional multivariate longitudinal data" context.

General framework. The setting of this paper is such that we want to incorporate high-dimensional time-dependent (longitudinal) features measured with error in a survival model. Let us consider the usual survival analysis framework. Following Andersen et al. [1993], let non-negative random variables $T^\star$ and C stand for the times of the event of interest and censoring times respectively. The event of interest could be for instance survival time, re-hospitalization, relapse or disease progression. We then denote T the right-censored time and Δ the censoring indicator, defined as

$$T = T^\star \wedge C \quad \text{and} \quad \Delta = \mathbb{1}(\{T^\star \leq C\})$$

respectively, where $a \wedge b$ denotes the minimum between two numbers a and b, and $\mathbb{1}(\cdot)$ the indicator function taking the value 1 if the condition in $(\cdot)$ is satisfied and 0 otherwise.

Let X denotes the p-dimensional vector of time-independent features (e.g. patients characteristics, therapeutic strategy, or omics features recorded at the begining of a study), and let $Y(t) = \left(Y^1(t), \ldots, Y^L(t)\right)^\top \in \mathbb{R}^L$ denote the value of the L-dimensional longitudinal outcome at time point $t \geq 0$, with $L > 0$.

Heterogeneity of the population. An assumption of heterogeneity within the patient population is frequently relevant in medical research where several differing profiles of subjects are expected [Bussy et al., 2019]. To take account of this, we introduce a latent variable $G \in \{0, \ldots, K-1\}$ modeling the $K \geq 1$ subgroups of different risk. Let us denote

$$\pi_{\xi_k}(x) = \mathbb{P}[G = k | X = x]$$

the latent class membership probability given time-independent features $x \in \mathbb{R}^p$, and consider a softmax link function given by

$$\pi_{\xi_k}(x) = \frac{e^{x^\top \xi_k}}{\sum_{k=0}^{K-1} e^{x^\top \xi_k}}$$

where $\xi_k \in \mathbb{R}^p$ denotes a vector of coefficients that quantifies the impact of each time-independent features on the probability that a subject belongs to the k-th group. From now on, all computations are done conditionally on features x.

Then, the idea is to allow flexibility in modeling the dependency between the longitudinal features and the event time. Joint models have the ability to produce prognostic predictions that can be dynamically updated according to the observed trajectory of the features, thus offering a powerful tool for clinical decision making in patient monitoring. But those questions are never considered in a high-dimensional framework [Proust-Lima and Taylor, 2009, Rizopoulos, 2011].

When validating predictions or comparing performances of competing models in a survival context, one can be interested either in the discriminative power of the predictive rule and use concordance measures such as the C-index [Heagerty and Zheng, 2005], or in the predictive accuracy of the rule [Schemper and Henderson, 2000]. Developments in the field of joint modeling have primarily focused on modeling and estimation, and most studies do not consider goodness of fit nor generalisation power [REF!]. However, practitioners will naturally require predictive prognostic tools to evaluate model fits and compare models.

Once the model is train, posterior risk classification can be made through posterior estimates of the latent class membership probabilities, and latent class model may be particularly suited for prediction problems. We propose to use this probabilities as discriminative marker rule for risk prediction and use it in the cross-validation procedure for selecting the best regularization hyper-parameters. Indeed, it makes sense from a practical point of view, since one want to obtain prognostic predictions as well as interpretations of the high-dimensional longitudinal features.

Main contribution. In this paper, we propose a method called *lights* (generaLized joInt hiGH-dimensional longiTudinal Survival) that `blabla... ajout description courte des caractéristiques du modèle une fois les maths terminées`.

Organization of the paper. A precise description of the model is given in Section 2. Section 3 focuses on a regularized version of the model to exploit dimension reduction and prevent overfitting. Inference is presented under this framework, as well as the convergence properties of the developed algorithm. Section 4 presents the simulation procedure used to evaluate the performance of our method and compares it with state-of-the-art ones. In Section 5, we apply our method to high-dimensional publicly available datasets. Finally, we discuss the obtained results in Section 6.

Notations. Throughout the paper, for every $q > 0$, we denote by $\|v\|_q$ the usual ℓ_q-quasi norm of a vector $v \in \mathbb{R}^m$, namely $\|v\|_q = (\sum_{k=1}^m |v_k|^q)^{1/q}$. We also denote $\|v\|_0 = |\{k : v_k \neq 0\}|$, where $|A|$ stands for the cardinality of a finite set A. For $u, v \in \mathbb{R}^m$, we denote by $u \odot v$ the Hadamard product $u \odot v = (u_1 v_1, \ldots, u_m v_m)^\top$. For a squared matrix M, vech(M) stacks columns of M one under another in a single vector, starting each column at its diagonal element. Finally, we write, for short, $\mathbf{1}_m$ (resp. $\mathbf{0}_m$) for the vector of $\mathbb{R}^m$ having all coordinates equal to one (resp. zero).

2 Method

In this section, we descibe the longitudinal and time-to-event submodels, as well as the required hypothesis in order to write a likelihood and draw inference for the lights model.

Group-specific marker trajectories. We suppose a group-specific marker trajectory and a generalized linear mixed model for each longitudinal marker given subgroup G, so that for the l-th outcome at time $t \geq 0$ one has

$$h_l\big(\mathbb{E}[Y^l(t)|b^l, G = k]\big) = m_k^l(t) \tag{1}$$

where h_l denotes a known one-to-one link function, and m_k^l the linear predictor such that

$$m_k^l(t) = u^l(t)^\top \beta_k^l + v^l(t)^\top b^l$$

where $u^l(t) \in \mathbb{R}^{p_l}$ is a row vector of (possibly) time-varying features with corresponding unknown fixed effect parameters β_k^l, and $v^l(t) \in \mathbb{R}^{r_l}$ is a row vector of (possibly) time-varying features with $r_l \leq p_l$ and where the corresponding subject-and-longitudinal outcome specific random effects b^l. A suitable distributional assumption for the random effects component is a zero-mean multivariate normal distribution [Hickey et al., 2016], that is

$$b^l \sim \mathcal{N}(0, D_{ll})$$

with $D_{ll} \in \mathbb{R}^{r_l \times r_l}$ the unstructured variance-covariance matrix that does not depend on the group membership. To account for dependence between the different longitudinal outcome types, we let $\text{Cov}[b^l, b^{l'}] = D_{ll'}$ for $l \neq l'$ and we denote

$$D = \begin{bmatrix} D_{11} & \cdots & D_{1L} \\ \vdots & \ddots & \vdots \\ D_{1L}^\top & \cdots & D_{LL} \end{bmatrix}$$

the global variance-covariance matrix. In the sequel of the paper, our aim is to associate the true and unobserved value $m_k^l(t)$ of the l-th longitudinal outcome at time t with the event outcome $T^\star$.

Group-specific risk of event. To quantify the effect of the longitudinal outcomes on the risk for an event, we use a relative risk model of the form

$$\lambda(t|\mathcal{M}_k(t), G = k, w) = \lambda_0(t) \exp\left\{ w^\top \gamma_{k,0} + \sum_{l=1}^{L} \sum_{a=1}^{\mathcal{A}} \gamma_{k,a}^l \varphi_a(t, \beta_k^l, b^l) \right\}, \tag{2}$$

where $\lambda_0(\cdot)$ denotes an unspecified baseline hazard function and given that $G = k$, $\mathcal{M}_k(t) = \{m_k^l(u), 0 \leq u < t\}$ denotes the history of the true unobserved longitudinal process up to time t, w is a q-vector of time-independant features with corresponding fixed effects $\gamma_k \in \mathbb{R}^q$, and $\gamma_{k,a}^l \in \mathbb{R}$ is the joint model association parameter corresponding to a known functional φ_a defining a shared association. Details on the choice of $(\varphi_a)_{a \in \mathcal{A}}$ are given in Section 4.1. From now on, all computations are done conditionally on features w.

Likelihood. Consider an independent and identically distributed (i.i.d.) cohort of n subjects

$$\mathcal{D} = \left\{ (x_1, y_1^1, \ldots, y_1^L, t_1, \delta_1), \ldots, (x_n, y_n^1, \ldots, y_n^L, t_n, \delta_n) \right\}$$

where for each subject $i = 1, \ldots, n$, process Y_i^l is measured n_i^l times at $t_{i1}^l, \ldots, t_{in_i^l}^l$ (which can differ between subjects and outcomes) with $t_{ij}^l \leq t_{ij+1}^l$ for all $j = 1, \ldots, n_i^l - 1$ and such that

$$y_i^l = (y_{i1}^l, \ldots, y_{in_i^l}^l)^\top \in \mathbb{R}^{n_i^l} \quad \text{with} \quad y_{ij}^l = Y_i^l(t_{ij}^l)$$

for all $l = 1, \ldots, L$. For the i-th subject, let us denote

$$\begin{cases} y_i &= (y_i^{1\top}, \ldots, y_i^{L\top})^\top \in \mathbb{R}^{n_i}, \\ b_i &= (b_i^{1\top}, \ldots, b_i^{L\top})^\top \in \mathbb{R}^{r}, \end{cases}$$

with $n_i = \sum_{l=1}^{L} n_i^l$ and $r = \sum_{l=1}^{L} r_l$ the total number of longitudinal measurements (for subject i) and the total dimension of the random effects respectively, as well as the following

design matrices

$$U_i = \begin{bmatrix} U_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_{iL} \end{bmatrix} \in \mathbb{R}^{n_i \times p} \quad \text{and} \quad V_i = \begin{bmatrix} V_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V_{iL} \end{bmatrix} \in \mathbb{R}^{n_i \times r}$$

where for all $l = 1, \ldots, L$, one writes

$$\begin{cases} U_{il} & = [u_i^l(t_{i1}^l) \cdots u_i^l(t_{in_i^l}^l)]^\top \in \mathbb{R}^{n_i^l \times p_l}, \\ V_{il} & = [v_i^l(t_{i1}^l) \cdots v_i^l(t_{in_i^l}^l)]^\top \in \mathbb{R}^{n_i^l \times r_l}. \end{cases}$$

Note that from now on, all computations are done conditionally on the design matrices $(U_i)_{i=1,\ldots,n}$ and $(V_i)_{i=1,\ldots,n}$. Suppose that conditional on the random effects, each Y_i^l is independent and the censoring times T_i are independent of b_i; both are standard modelling assumptions. For all $k = 0, \ldots, K-1$, we denote

$$\beta_k = (\beta_k^{1\top}, \ldots, \beta_k^{L\top}) \in \mathbb{R}^p$$

with $p = \sum_{l=1}^L p_l$ and

$$M_{ik} = U_i \beta_k + V_i b_i \in \mathbb{R}^{n_i}.$$

Given (1), each y_i^l is assumed to be from a one-parameter exponential family with respect to a reference measure which is either the Lebesgue measure (e.g. in the Gaussian case) or the counting measure (e.g. in the logistic cases). The conditional distribution of $y_i | b_i, G_i = k$ is then assumed to be from a distribution with a density of the form

$$f(y_i | b_i, G_i = k) = \exp\left\{ (y_i \odot \Phi_i)^\top M_{ik} - c_\phi(M_{ik}) + d_\phi(y_i) \right\}, \tag{3}$$

with

$$\Phi_i = [\phi_1^{-1} \mathbf{1}_{n_i^1}^\top \cdots \phi_L^{-1} \mathbf{1}_{n_i^L}^\top]^\top \in \mathbb{R}^{n_i}$$

and $\phi = (\phi_1, \ldots, \phi_L)^\top \in \mathbb{R}^L$. The density described in (3) encompasses several distributions, see Table 1. The functions $c_\phi(\cdot)$ and $d_\phi(\cdot)$ are known as well as the dispersion parameters ϕ_l, while parameters β_k have to be estimated. Note that $d_\phi(\cdot)$ is related to the normalizing constant.

Model	Support	Use cases	ϕ_l	$h_l(\cdot)$ in (1)	$c_{\phi_l}(\cdot)$
Normal	$\mathbb{R}$	Continuous response data	σ_l^2	$z \mapsto z$	$z \mapsto \phi_l^{-1} \frac{z^2}{2}$
Categorical	$\{0,1\}^C$	Outcome with C modalities	1	$z \mapsto \log\left(\frac{z}{1-z}\right)$	$z \mapsto \phi_l^{-1} \frac{e^z}{1+e^z}$
Poisson	$\mathbb{N}$	Count of occurrences	1	$z \mapsto e^z$	$z \mapsto \phi_l^{-1} e^z$

Table 1: Examples of standard distributions that fit in the considered setting.

Let us denote

$$\theta = (\xi_0, \ldots, \xi_{K-1}, \beta_0, \ldots, \beta_{K-1}, \phi^\top, \text{vech}(D), \lambda_0(t), \gamma_0, \ldots, \gamma_{K-1})^\top$$

the collection of unknown parameters that we want to estimate, with $\gamma_k = $ `to be defined properly`. To write the log-likelihood $\ell_n(\theta)$ (rescaled by n^{-1}) for samples in $\mathcal{D}$, corresponding to the joint distribution of the time-to-event and longitudinal outcomes, let us assume

that the random effects vector b_i and the group membership account for the association between the longitudinal and event outcomes, that is

$$f(t_i, \delta_i, y_i | b_i, G_i = k; \theta) = f(t_i, \delta_i | b_i, G_i = k; \theta) f(y_i | b_i, G_i = k; \theta) \tag{4}$$

for all $i = 1, \ldots, n$. Hypothesis (4) is `faire le lien avec les hyps classiques de la litterature` [REFs]. Then, one has

$$\begin{aligned}
\ell_n(\theta) &= \ell_n(\theta; \mathcal{D}) \\
&= n^{-1} \sum_{i=1}^{n} \log \int_{-\infty}^{+\infty} \sum_{k=0}^{K-1} \pi_{\xi_k}(x_i) f(t_i, \delta_i | b_i, G_i = k; \theta) f(y_i | b_i, G_i = k; \theta) f(b_i; \theta) \mathrm{d}b_i,
\end{aligned} \tag{5}$$

where

$$f(t_i, \delta_i | b_i, G_i = k; \theta) = \left[\lambda(t_i | \mathcal{M}_k(t_i), G_i = k, w_i) \right]^{\delta_i} \exp \left\{ -\int_0^{t_i} \lambda(u | \mathcal{M}_k(u), G_i = k, w_i) \mathrm{d}u \right\}$$

and

$$f(b_i; \theta) = (2\pi)^{-\frac{r}{2}} |D|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} b_i^\top D^{-1} b_i \right\}.$$

3 Inference

In order to avoid overfitting and to improve the prediction power of the lights model, we use Elastic-Net regularization [Zou and Hastie, 2005] by minimizing the penalized objective

$$\ell_n^{\mathrm{pen}}(\theta) = -\ell_n(\theta) + \sum_{k=0}^{K-1} \zeta_{1,k} \|\xi_k\|_{\mathrm{E\text{-}N}, \eta} + \zeta_{2,k} \|\beta_k\|_{\mathrm{E\text{-}N}, \eta} + \zeta_{3,k} \|\gamma_k\|_{\mathrm{E\text{-}N}, \eta} \tag{6}$$

where for all $k = 0, \ldots, K - 1$, we add a linear combination of the lasso (ℓ_1) and ridge (squared ℓ_2) penalties of the vectors ξ_k, β_k and γ_k for tuning hyperparameters $(\zeta_{1,k}, \zeta_{2,k}, \zeta_{3,k})^\top \in \mathbb{R}_+^3$, with a fixed $\eta \in [0, 1]$ and

$$\|v\|_{\mathrm{E\text{-}N}, \eta} = (1 - \eta)\|v\|_1 + \frac{\eta}{2}\|v\|_2^2$$

for any vector v. `expliquer interpretation de` $(\zeta_{1,k}, \zeta_{2,k}, \zeta_{3,k})$

One advantage of this regularization method is its ability to perform model selection (the lasso part) and pinpoint the most important features relatively to the prediction objective. On the other hand, the ridge part allows to handle potential correlation between time-independent features or between longitudinal outcomes. Note that in practice, the intercept is not regularized.

3.1 A Fast Stochastic Approximation of a Quasi-Newton EM

`dérouler l'algo. idée: QNEM +` Karimi et al. [2020]

3.2 Convergence to a stationary point

`théorème avec étude de convergence`

3.3 Cross-validation procedure

Posterior group membership probabilities. Les us denote

$$\hat{\pi}_{ik}^{Y,T} = \mathbb{P}_{\hat{\theta}}[G_i = k | y_i, t_i, \delta_i] \tag{7}$$

utiliser ce marker pour cross-val les $(\zeta_{1,k}, \zeta_{2,k}, \zeta_{3,k})$ avec du C-index + use randomized grid search [Bergstra and Bengio, 2012]

4 Performance evaluation

In this section, we first give numerical details and briefly introduce the models we consider for performance comparisons. Then, we provide details regarding the simulation study and data generation. The chosen metrics for evaluating performances are then presented, followed by the results.

4.1 Numerical details

Specification of the design matrices. In many practical applications, subjects show highly nonlinear longitudinal trajectories. In (5), the complete longitudinal history is required for the computation of the survival function. Hence, in order to produce a good estimate of $\mathcal{M}(t)$, we consider a flexible representations for $u^l(t)$ and $v^l(t)$ using a high-dimensional vector of functions of time t expressed in terms of splines.

Specification of the functionals $(\varphi_a)_{a \in \mathcal{A}}$**.**

4.2 Competing models

State-of-the-art implementations in `R`:

- recent package in multivariate settings: `joineRML` [Hickey et al., 2018]
- reference one: `JM` [Rizopoulos, 2010]
- MCMC methods: `JMbayes` [Rizopoulos, 2014] (voir slide 132 http://www.drizopoulos.com/courses/Int/JMwithR_CEN-ISBS_2017.pdf)
- `joineR` [Philipson et al., 2018]
- classical time-dependent Cox model: `survival`, explanations on how to do this in Zhang et al. [2018]

> still nothing in `Python` ! I'll propose an implementation of the model in `Python` (today paramount in machine learning, it will be an additional asset of the paper).

4.3 Simulation design

In order to assess the proposed method, we perform an extensive Monte Carlo simulation study.

simulation of time-varying features in a Cox model: see Therneau et al. [2017]

4.4 Metric

We detail in this section the metric considered to evaluate risk prediction performances. Let us denote by M the marker under study. We denote by h the probability density function of marker M, and assume that the marker is measured once at $t = 0$.

A common concordance measure that does not depend on time is the C-index [Harrell et al., 1996] defined by

$$\mathcal{C} = \mathbb{P}[M_i > M_j | T_i^\star < T_j^\star],$$

with $i \neq j$ two independent subjects (which does not depend on i, j under the $i.i.d.$ sample hypothesis). In our case, $T^\star$ is subject to right censoring, so one would typically consider the modified $\mathcal{C}_\tau$ defined by

$$\mathcal{C}_\tau = \mathbb{P}[M_i > M_j | T_i < T_j, T_i < \tau],$$

with τ corresponding to the fixed and prespecified follow-up period duration [Heagerty and Zheng, 2005]. A Kaplan-Meier estimator for the censoring distribution leads to a nonparametric and consistent estimator of $\mathcal{C}_\tau$ [Uno et al., 2011], already implemented in the R package `survival`. Hence in the following, we consider the C-index metric to assess performances.

4.5 Results of simulation

Let us present now the simulation results.

5 Applications

```
PBC dataset: open source, multidim mais dimension ~15 (voir https://rdrr.io/
cran/JM/man/pbc.html).
essayer de trouver 1 dataset en + grande dimension?

    other idea ++ use MIMIC II or III dataset !
```
https://www.nature.com/articles/sdata201635/
https://ieeexplore.ieee.org/abstract/document/1166854
https://www.sciencedirect.com/science/article/abs/pii/S2213260014702395
https://www.sciencedirect.com/science/article/pii/S2352556818302169

6 Conclusion

Software

All the methodology discussed in this paper is implemented in `Python` and `R`. The code is available from https://github.com/SimonBussy/? in the form of annotated programs, together with a notebook tutorial.

Acknowledgements

Conflict of Interest: None declared.

Appendices

A   Proofs

References

P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding. *Statistical models based on counting processes.* Springer Series in Statistics. Springer-Verlag, New York, 1993.

James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305, 2012.

Simon Bussy, Agathe Guilloux, Stéphane Gaïffas, and Anne-Sophie Jannot. C-mix: A high-dimensional mixture model for censored durations, with applications to genetic data. *Statistical methods in medical research*, 28(5):1523–1539, 2019.

Frank E Harrell, Kerry L Lee, and Daniel B Mark. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*, 15:361–387, 1996.

Patrick J Heagerty and Yingye Zheng. Survival model predictive accuracy and roc curves. *Biometrics*, 61(1):92–105, 2005.

Graeme L Hickey, Pete Philipson, Andrea Jorgensen, and Ruwanthi Kolamunnage-Dona. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC medical research methodology*, 16(1):117, 2016.

Graeme L Hickey, Pete Philipson, Andrea Jorgensen, and Ruwanthi Kolamunnage-Dona. joinerml: a joint model and software package for time-to-event and multivariate longitudinal outcomes. *BMC medical research methodology*, 18(1):50, 2018.

Miran A Jaffa, Mulugeta Gebregziabher, and Ayad A Jaffa. A joint modeling approach for right censored high dimensional multivariate longitudinal data. *Journal of biometrics & biostatistics*, 5(4), 2014.

Belhal Karimi, Marc Lavielle, and Eric Moulines. f-saem: A fast stochastic approximation of the em algorithm for nonlinear mixed effects models. *Computational Statistics & Data Analysis*, 141:123–138, 2020.

Haiqun Lin, Charles E McCulloch, and Susan T Mayne. Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Statistics in Medicine*, 21(16):2369–2382, 2002.

Pete Philipson, Ines Sousa, Peter J Diggle, and Paula Williamson. Package 'joiner'. 2018.

Cécile Proust-Lima and Jeremy MG Taylor. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modeling approach. *Biostatistics*, 10(3):535–549, 2009.

Dimitris Rizopoulos. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3):819–829, 2011.

Dimitris Rizopoulos. The r package jmbayes for fitting joint models for longitudinal and time-to-event data using mcmc. *arXiv preprint arXiv:1404.7625*, 2014.

Dimitris D Rizopoulos. Jm: An r package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software (Online)*, 35(9):1–33, 2010.

Michael Schemper and Robin Henderson. Predictive accuracy and explained variation in cox regression. *Biometrics*, 56(1):249–255, 2000.

Terry Therneau, Cindy Crowson, and Elizabeth Atkinson. Using time dependent covariates and time dependent coefficients in the cox model. *Survival Vignettes*, 2017.

Anastasios A Tsiatis and Marie Davidian. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, pages 809–834, 2004.

Hajime Uno, Tianxi Cai, Michael J Pencina, Ralph B D'Agostino, and LJ Wei. On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in medicine*, 30(10):1105–1117, 2011.

Ping Wang, Wei Shen, and Mark Ernest Boye. Joint modeling of longitudinal outcomes and survival using latent growth modeling approach in a mesothelioma trial. *Health Services and Outcomes Research Methodology*, 12(2-3):182–199, 2012.

Michael S Wulfsohn and Anastasios A Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, pages 330–339, 1997.

Menggang Yu, Ngayee J Law, Jeremy MG Taylor, and Howard M Sandler. Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, pages 835–862, 2004.

Zhongheng Zhang, Jaakko Reinikainen, Kazeem Adedayo Adeleke, Marcel E Pieterse, and Catharina GM Groothuis-Oudshoorn. Time-varying covariates and coefficients in cox regression models. *Annals of translational medicine*, 6(7), 2018.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.