MOX-Report No. 27/2020

Modelling dynamic covariates effect on survival via Functional Data Analysis: application to the MRC BO06 trial in osteosarcoma

Spreafico, M.; Ieva, F.; Fiocco, M.

Modelling dynamic covariates effect on survival via Functional Data Analysis: application to the MRC BO06 trial in osteosarcoma

Marta Spreafico[1,2,3]   Francesca Ieva[1,3,4]   Marta Fiocco[2,5,6]

marta.spreafico@polimi.it   francesca.ieva@polimi.it   m.fiocco@math.leidenuniv.nl

[1]MOX – Department of Mathematics, Politecnico di Milano, Milan 20133, Italy

[2]Mathematical Institute, Leiden University, Leiden, The Netherlands

[3]CHRP, National Center for Healthcare Research and Pharmacoepidemiology, Milan 20126, Italy

[4]CADS, Center for Analysis Decisions and Society, Human Technopole, Milan 20157, Italy

[5]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

[6]Trial and Data Center, Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands

April 22, 2020

Abstract

Time-varying covariates are of great interest in clinical research since they represent dynamic patterns which reflect disease progression. In cancer studies biomarkers values change as functions of time and chemotherapy treatment is modified by delaying a course or reducing the dose intensity, according to patient's toxicity levels. Models for time-to-event to deal with the dynamic nature of time-varying covariates during follow-up are necessary, still not well developed. In this work, innovative methods to represent time-dependent covariates by means of Functional Data Analysis (FDA) and how to include them into survival models are discussed. This new approach was applied to osteosarcoma data from the MRC BO06/EORTC 80931 randomized clinical trial, new insights into the clinical research. Time-varying covariates related to alkaline phosphatase (ALP) and chemotherapy dose during treatment were considered. Processes dynamics over time were investigated and additional information that may be related to the survival were included into the time-to-event models. High ALP levels reflected poor overall survival. Although dose-intense profiles were not associated with a better survival, the strength of our method is the ability to detect differences between patients with different biomarker evolution and treatment response, even when randomised to the same regimen. This aspect is seldom addressed in the literature.

Key-words: Time-varying covariates; Survival analysis; Functional data analysis; Joint models; Osteosarcoma

1 Introduction

Osteosarcoma is a malignant bone tumour mainly affecting children and young adults. Although osteosarcoma is the most common primary malignant bone cancer, it is a rare disease and has an annual incidence of 3-4 patients per million[1]. Multidisciplinary management including neoadjuvant and adjuvant chemotherapy with aggressive surgical resection[2] or intensified chemotherapy[3] has improved clinical outcomes although the overall 5-year survival has remained unchanged in the last 40 years at 60–70%. Therefore, it is extremely important to provide an effective tool to evaluate the prognosis for osteosarcoma and to guide the diagnosis.

Time-varying (or time-dependent) covariates are often of interest in clinical and epidemiological research: patients are followed during the study and subject-specific measurements are recorded at each visit. Well-known examples include biomarkers which change during follow-up or cumulative exposure to medications,[4] such as chemotherapy. Depending on patients' treatment history or development of toxicity, biomarker values change and chemotherapy treatment is modified by delaying a course or reducing the dose intensity. Hence to study the association between time-varying responses with a time-to-event outcome (e.g. death) is a challenging task which could offer new insights into the direction of personalised treatment.

For osteosarcoma, alkaline phosphatase (ALP) has been identified as prognostic marker.[5] High ALP levels usually indicate a bad liver or bone condition.[6] The role of ALP as tumour marker for osteosarcoma has not been established, although several studies suggested that high ALP level is associated with poor overall or event-free survival and presence of metastasis.[7,8] In previous studies, ALP levels have always been incorporated in survival models as baseline time-fixed covariate. Chemotherapy is modelled by different allocated regimens, without considering changes in drug assumptions, i.e. delays or dose reduction, over time.[9] It had been shown that there is mismatch between target and achieved dose of chemotherapy[10] and the impact on patients' survival is still unclear. Therefore, the most appropriate way to look at both received chemotherapy dose and biomarkers is to model them as time-varying variables. In this way it is possible to investigate the dynamics over time and to use additional information that may be related to survival. This approach represents a novelty for osteosarcoma treatment and the application is of great interest for statistical modelling.

Models for time-to-event able to deal with the dynamic nature of time-varying responses during follow-up are not well developed and effects on survival are still unclear. During the past two decades research into joint modelling for longitudinal and time-to-event data has received a lot of attention.[11–19] This approach allows to jointly represent longitudinal sub-models for internal time-varying processes generated by the subject under study and an event sub-model for the time-to-event outcome of interest. When the outcome processes are correlated, joint modelling reduce biases in the effect size estimates, improve efficiency and prediction and can be applicable to outcome surrogacy.[18] Functional Data Analysis (FDA) has been increasingly used to analyse, model and predict dynamic processes.[20–27] The idea behind FDA and functional models is to express discrete observations arising from time series, i.e. longitudinal time-varying observations, in the form of functions.[21,22] The key point of a functional representation is that it is able to incorporate trends and variations in the evolution of the process over time.[24] Additional information implied by changes in values and smoothness of functions during follow up can therefore be investigated. Since functional data are infinite-dimensional covariates, some dimension reduction methods to summarize and select a finite dimensional set of elements representing the most important features of each covariate have to be used. This information can then be included into time-to-event models. To model the relationship between survival outcomes and a set of finite and infinite dimensional predictors Functional Linear Cox Regression Models (FLCRM) have been recently proposed.[28–31] Kong *et al.*[31] characterized the joint effects of both functional and

scalar predictors on time-to-event outcome employing Functional Principal Component Analysis (FPCA). FPCA is one of the most popular dimension reduction method in FDA and it is used to summarise each function to a finite set of covariates through FPCs scores, while loosing a minimum part of the information. FDA provides a novel modelling and prediction approach, with the potential for many applications across a range of public health and biomedical applications.[24]

Motivated by a clinical question concerning the effect of actually received chemotherapy doses on survival for osteosarcoma patients, innovative methods to represent time-varying ALP biomarker and chemotherapy dose during treatment are proposed. In joint models, a longitudinal sub-model for the time-varying process can be specified only if the covariate is internal (or endogenous),[32] i.e. if the covariate is the output of a stochastic process that is generated by the individual under study. ALP biomarker is an example of internal time-dependent covariate, since it changes according to patient's health status, whereas chemotherapy dose is external (or exogenous), since its value is fixed at the beginning or changes according to the guidelines provided in the protocol. First, joint modelling to investigate how patients' survival are affected by endogenous ALP longitudinal values is applied. Then, FDA techniques are used to represent both internal time-varying biomarker and external chemotherapy dose in terms of functions, in order to study their effects on survival. Additional information contained into the evolution of the functions over time can be included into a survival model, extending FLCRM to the case of multiple functional predictors. Within this work three important novelties are proposed: (i) application of advanced statistical techniques able to deal with time-varying covariates in the field of osteosarcoma treatment; (ii) reconstruction of the functional representations for ALP biomarker and chemotherapy dose, able to incorporate trends and variations in the evolution of the processes over time, starting from clinical database; (iii) extension of FLCRM to the multivariate case. This approach can provide more information about the effect of individualize treatment adaption on survival for osteosarcoma patients.

The remaining part of the paper is organized as follows. In Section 2 the statistical methodologies are introduced. MRC BO06/EORTC 80931 data related to a randomized controlled trial for patients with osteosarcoma,[3] with data inclusion criteria, trial protocol and longitudinal representations of time-varying covariates, are described in Section 3. Results for MRC BO06 data are presented in Section 4. Section 5 ends with a discussion of strengths and limitations of the current approaches, identifying some developments for future research in the field.

2 Statistical methods

In this section two approaches for modelling time-to-event data, Overall Survival OS, in presence of time-varying covariates are introduced. In Section 2.2 the joint model proposed by Rizopoulos[16,33] for dealing with time-to-event and endogenous longitudinal covariates is briefly recalled. In Section 2.3 the FLCRM[31] is extended to multiple functional predictors. Statistical analysis is performed in the R-software environment.[34] R code was developed by the first author and is available on request.

2.1 Time-varying covariates and survival frameworks

A time-varying (or time-dependent) covariate, $X(t)$, is a covariate whose value can change over the duration of follow-up (e.g., time-varying biomarkers, current use of medication, and cumulative dose of drugs). Kalbfieish and Prentice[32] defined different types of time-varying covariates: *external* (exogenous) or *internal* (endogenous) covariates. External covariates are unaffected by the process and their values over time are defined or established from the beginning, prior to the commencement of the study. An example is the cumulative dose of chemotherapy in clinical
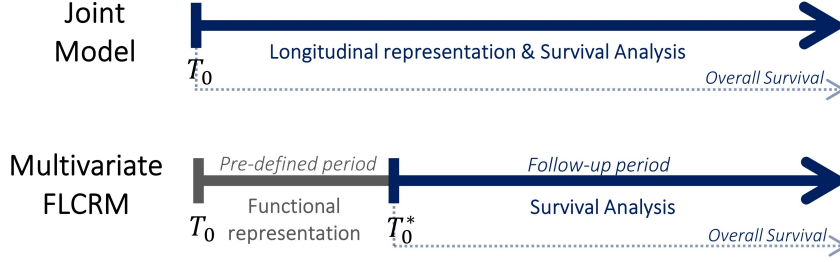
Figure 1: Time-varying representation and Overall Survival (OS) for different models.

trials, where the dose is changed by clinicians according to a previously defined mechanism. In contrast, internal covariates are "the output of a stochastic process generated by the individual under study", i.e. they are related to the behaviour of the individuals over time and their time paths are jointly determined with the responses of interest. According to these definitions, ALP biomarker is an example of internal time-dependent covariate since it changes according to the patient's health status, whereas chemotherapy dose is an external covariate since its value is fixed at the beginning or changes according to the guidelines provided in the protocol.

In time-to-event analysis, for each patient $i \in \{1, ..., N\}$ the observed data is (T_i, δ_i), where $T_i = \min(T_i^*, C_i)$ denotes the observed event time, T_i^* is the true event time, C_i is the censoring time and $\delta_i = I(T_i^* \geq C_i)$ is the event indicator, with $I(\cdot)$ being the indicator function that takes the value 1 when $T_i^* \geq C_i$, and 0 otherwise. In this study, the event of interest is death for any cause.

To model time-varying covariates into a time-to-event framework the internal or external nature of the covariates has to be considered. Using the joint modelling approach, only internal time-varying covariates can be jointly modelled with the time-to-event outcome, since a longitudinal sub-model can be used only if the dynamic covariate is the output of a stochastic process generated by the individual under study. In this case, the Overall Survival (OS) is measured from the date of randomization (T_0 in Figure 1) to the date of death or last follow-up date. The multivariate FLCRM is able to manage both internal and external covariates. A *pre-defined period* starting from the date of randomization (T_0) is used to reconstruct and summarize the functional representations of time-varying covariates (internal or external). Then, survival analysis is performed over the *follow-up period* and Overall Survival (OS) is measured from the end of the *pre-defined period* (T_0^* in Figure 1) to the date of death or last follow-up date. For this approach only patients still alive at T_0^* are incorporated in the survival analysis. This choice could imply a loss of information in case of exclusion of too many early dying patients.

In this application only one patient died before T_0^*. First joint model to simultaneously study internal ALP biomarker and time-to-death is applied (Section 4.1). Then, both internal ALP biomarker and external chemotherapy dose as functional predictors are considered and a multivariate FLCRM in order to assess how both time-varying covariates affect patients' survival is estimated (Section 4.2).

2.2 Joint model for time-to-event and internal longitudinal data

Let $\mathcal{D}_N = \{T_i, \delta_i, \mathbf{y}_i; i = 1, ..., N\}$ denote a sample from the target population, where $\mathbf{y}_i$ is the $n_i \times 1$ longitudinal response vector for the i-th subject. The term y_{il} denotes the value of the longitudinal process taken at time t_{il} ($l = 1, ..., n_i$). The general form of joint models is as

follows:

$$g[\mathsf{E}\{y_i(t)|\mathbf{b}_i\}] = \eta_i(t) = \mathbf{x}_i^T(t)\boldsymbol{\beta} + \mathbf{z}_i^T(t)\mathbf{b}_i \tag{1}$$

$$h_i(t|\mathcal{H}_i(t),\boldsymbol{\omega}_i) = h_0(t)\exp\left\{\boldsymbol{\gamma}^T\boldsymbol{\omega}_i + f(\mathcal{H}_i(t),\mathbf{b}_i,\boldsymbol{\alpha})\right\}, \qquad t > 0. \tag{2}$$

The longitudinal process (1) is a generalized linear mixed effects model where $g(\cdot)$ denotes a known one-to-one monotonic link function, $y_i(t)$ denotes the value of the longitudinal process for the i-th subject at time point t, $\boldsymbol{\beta}$ is the vector of the unknown fixed effects parameters, $\mathbf{b}_i$ is the vector of subject-specific random effects, $\mathbf{x}_i(t)$ and $\mathbf{z}_i(t)$ denote the time-dependent vectors for the fixed and random effect, respectively. The event process (2) assumes that the risk $h_i(\cdot)$ for an event depends on a function $f(\cdot)$ of the subject-specific linear predictor $\eta_i(t)$; $\mathcal{H}_i(t) = \{\eta_i(s), 0 \leq s < t\}$ represents the history of the underlying longitudinal process up to time point t, $h_0(\cdot)$ denotes the baseline hazard function, $\boldsymbol{\omega}_i$ is a vector of baseline covariates with corresponding regression coefficient $\boldsymbol{\gamma}$. The parameter $\boldsymbol{\alpha}$ is the vector that quantifies the association between the marker process up to time t and the hazard of an event at the same time point. The baseline hazard function $h_0(\cdot)$ is modelled using a B-splines approach. In `JMbayes` package,[33] the estimation of the model parameters is based on a Bayesian approach, using MCMC algorithms. Details regarding Bayesian estimation of joint models can be found in Ibrahim *et al.*[35] and papers by Brown *et al.*[36] and Rizopoulos.[33] In Section 4.1, this framework will be used to jointly model internal ALP biomarker and time-to-death.

2.3 Functional linear Cox regression model with multiple functional predictors

The model introduced in Kong *et al.*[31] is extended to multiple functional predictors, in order to provide a suitable framework for the application of interest.

Let $\tilde{X}_i^{(1)}, ..., \tilde{X}_i^{(M)}$ be a set of M functional predictors for individual i. The functional covariates are included in the Cox model[37] as:

$$h_i\left(t|\boldsymbol{\omega}_i, \tilde{x}_i^{(1)}, ..., \tilde{x}_i^{(M)}\right) = h_0(t)\exp\left\{\boldsymbol{\gamma}^T\boldsymbol{\omega}_i + \sum_{m=1}^{M}\int_{S_m}\tilde{x}_i^{(m)}(s)\alpha^{(m)}(s)ds\right\} \tag{3}$$

where $h_0(t)$ is the baseline hazard function, $\boldsymbol{\omega}_i$ is the vector of scalar (non functional) covariates with regression parameters $\boldsymbol{\gamma}$. The vector $\left(\tilde{x}_i^{(1)}, ..., \tilde{x}_i^{(M)}\right)$ is the realization of the M-variate functional data for the i-th individual; S_m and $\alpha^{(m)}(s)$ are compact sets of $\mathbb{R}$ and the functional regression parameters respectively.

By applying FPCA, each functional trajectory $\tilde{x}_i^{(m)}(s)$ can be approximated with a finite sum of K_m orthonormal basis $\left\{\xi_1^{(m)}, ..., \xi_{K_m}^{(m)}\right\}$:

$$\tilde{x}_i^{(m)}(s) \approx \mu^{(m)}(s) + \sum_{k=1}^{K_m} f_{ik}^{(m)}\xi_k^{(m)}(s) \tag{4}$$

where $\mu^{(m)}(s)$ is the functional mean and $f_{ik}^{(m)}$ is the FPC score of individual i related to the k-th orthonormal base $\xi_k^{(m)}$. In the analyses K_m is chosen in such a way that the percentage of

explained variance is $\geq 95\%$. From (4) the integrals in (3) can be approximated considering:

$$\int_{S_m} \left[\tilde{x}_i^{(m)}(s) - \mu^{(m)}(s) \right] \alpha^{(m)}(s)ds \approx \int_{S_m} \sum_{k=1}^{K_m} f_{ik}^{(m)} \xi_k^{(m)}(s) \alpha^{(m)}(s)ds$$

$$= \sum_{k=1}^{K_m} f_{ik}^{(m)} \int_{S_m} \xi_k^{(m)}(s) \alpha^{(m)}(s)ds \qquad (5)$$

$$= \sum_{k=1}^{K_m} f_{ik}^{(m)} \alpha_k^{(m)}$$

where $\alpha_k^{(m)}$ is the scalar representing the quantity $\int_{S_m} \xi_k^{(m)}(s)\alpha^{(m)}(s)ds$. Introducing approximation (5) in Equation (3), the hazard function becomes:

$$h_i\left(t|\boldsymbol{\omega}_i, \tilde{x}_i^{(1)}, ..., \tilde{x}_i^{(M)}\right) = h_0(t)\exp\left\{\boldsymbol{\gamma}^T\boldsymbol{\omega}_i + \sum_{m=1}^{M}\left[\int_{S_m}\mu^{(m)}(s)\alpha^{(m)}(s)ds + \sum_{k=1}^{K_m} f_{ik}^{(m)}\alpha_k^{(m)}\right]\right\}$$

$$= h_0^*(t)\exp\left\{\boldsymbol{\gamma}^T\boldsymbol{\omega}_i + \sum_{m=1}^{M}\sum_{k=1}^{K_m} f_{ik}^{(m)}\alpha_k^{(m)}\right\}$$

(6)

where $\alpha_k^{(m)} = \int_{S_m} \xi_k^{(m)}(s)\alpha^{(m)}(s)ds$ and $h_0^*(t) = h_0(t)\exp\left\{\sum_{m=1}^{M}\int_{S_m}\mu^{(m)}(s)\alpha^{(m)}(s)ds\right\}$. Therefore, defining the following quantities:

$$\boldsymbol{\theta} = \left[\boldsymbol{\gamma}^T, \left(\alpha_1^{(1)}, ..., \alpha_{K_1}^{(1)}\right), ..., \left(\alpha_1^{(M)}, ..., \alpha_{K_M}^{(M)}\right)\right]^T$$

$$\mathbf{w}_i = \left[\boldsymbol{\omega}_i^T, \left(f_{i,1}^{(1)}, ..., f_{i,K_1}^{(1)}\right), ..., \left(f_{i,1}^{(M)}, ..., f_{i,K_M}^{(M)}\right)\right]^T$$

and substituting them in Equation (6), through FPCA the multivariate FLCRM can be expressed as Cox model:

$$h_i(t) = h_0(t)\exp\left\{\boldsymbol{\theta}^T\mathbf{w}_i\right\}. \qquad (7)$$

All the properties of the Cox model still hold in this framework and the vector of coefficients $\boldsymbol{\theta}$ can be estimated by maximising the partial likelihood function.[37]

In Section 4.2, this framework will be used to assess how both time-varying internal ALP biomarker and external chemotherapy dose affect patients' survival.

3 MRC BO06 randomized clinical trial data

In Section 3.1 the selected cohort of patients is illustrated. Patient characteristics are presented as numbers and percentages for categorical variables, and as medians with interquartile ranges (IQRs) for continuous variables (Table 1). In Section 3.2 the time-varying characteristics related to ALP biomarker and chemotherapy dose are introduced using a longitudinal representation.

3.1 Sample selection and baseline characteristics

Data from the MRC BO06/EORTC 80931 Randomized Clinical Trial (RCT) for patients with non-metastatic high-grade osteosarcoma recruited between 1993 and 2002[3] were analysed. Patients were randomized between conventional (Reg-C) and dose-intense (Reg-DI) regimens. Both
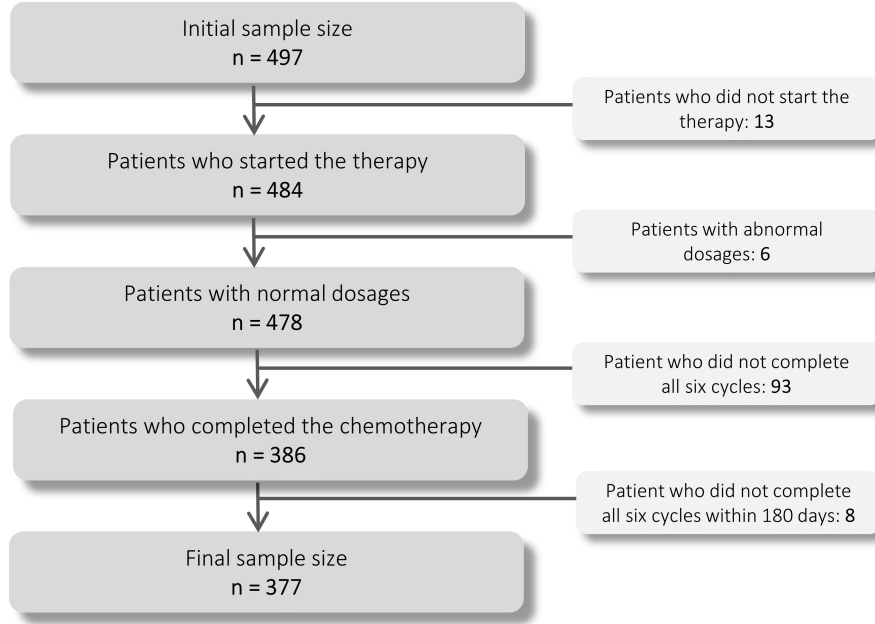
Figure 2: Flowchart of cohort selection.

arms have six cycles of the same course of chemotherapy, 75 mg/m^2 of doxorubicin (DOX) plus 100 mg/m^2 of cisplatin (CDDP), with different time schedule: in *Reg-DI* cycles are every two weeks, whereas in *Reg-C* they are every three weeks. Additional details concerning the protocol can be found in Appendix A.

The dataset included 497 eligible patients; 19 patients who did not start chemotherapy (13) or reported an abnormal dosage of one or both agents (6) were excluded. Motivated by the clinical research question concerning the effect of doses intensity on survival, only patients who completed all six cycles within 180 days were included in the analysis, while 93 patients who did not complete the therapy and 8 who did not terminated the last cycle within 180 days were excluded. The final cohort of 377 patients included in the analyses (75.9% of the initial sample) is shown in Figure 2.

Follow-up starts from date of randomization. Among 377 patients, 229 (60.7%) were males. At baseline (Table 1), median age was 15 years (51.2% were older than 15). The median value of baseline White Blood Count (WBC) was 7.65 (IQR = [6.30; 9.10]) [$\times 10^9$/L] and regimen Dose-Intense was allocate in 52.3% of the patients (197). Median follow-up time, computed using the reverse Kaplan-Meier method,[38] was 51.98 months (IQR = [28.65; 82.79]) and 297 patients (78.8%) was alive at the last follow-up visit.

3.2 Time-varying characteristics

In MRC BO06 RCT, the values of ALP biomarkers were measured through laboratory tests, usually performed before each cycle of chemotherapy. For each patient i the vector of longitudinal values of ALP measurements is given as $\mathbf{y}_i = \{y_i(t_{il}), l = 1, ..., n_i\}$, where t_{il} is the time of the l-th ALP test, $y_i(t_{il})$ is the value of ALP measurement at time t_{il} and n_i is the number of

Baseline characteristic	
Patients	377
Age [years]	
< 15	184 (48.8%)
≥ 15	193 (51.2%)
Median (IQR)	15 (11; 18)
Minimum/maximum	3/40
Gender	
Female	148 (39.3%)
Male	229 (60.7%)
Allocated treatment	
Regimen-C	180 (47.7%)
Regimen-DI	197 (52.3%)
White Blood Count [$\times 10^9$/L]	
Median (IQR)	7.65 (6.30; 9.10)
Minimum/maximum	3.60/16.20

Table 1: Patients' characteristics at baseline.

different laboratory tests. The left panel of Figure 3 shows the longitudinal trajectories of $\mathbf{y}_i$ over time. Each line represents the time-varying ALP biomarker for a specific patient coloured by event status (black: *Censored*, red: *Dead*).

The time-varying standardized cumulative dose of chemotherapy is now introduced. Let $j \in \{1, ..., 6\}$ be the cycle index and t_{ij} the time of the j-th cycle for the i-th patient. The standardized cumulative dose of chemotherapy (DOX+CDDP) for the i-th patient at time t_{ij} is defined as:

$$
\begin{aligned}
\delta_i(t_{ij}) &= \frac{\text{Cumulative dose of } DOX\text{+}CDDP \text{ until cycle } j \; [mg/m^2]}{\text{Total target dose at the end of six cycles } [mg/m^2]} \\
&= \frac{1}{175 \, [mg/m^2] \cdot 6} \cdot \sum_{k=1}^{j} \frac{DOX_{ik} + CDDP_{ik}}{surface\,area_{ik}} \left[\frac{mg}{m^2}\right].
\end{aligned}
\tag{8}
$$

This can be interpreted as the regulated Received Dose Intensity (rRDI) introduced by Lancia *et al.*[9] evaluated over real time and not over cumulative time on treatment. For each patient i, the vector of longitudinal values of standardized cumulative dose of chemotherapy over time is defined as $\boldsymbol{\delta}_i = \{\delta_i(t_{ij}), j = 1, ..., 6\}$. The right panel of Figure 3 shows the longitudinal trajectories $\boldsymbol{\delta}_i$ over time. Each line represents the individual time-varying standardized cumulative chemotherapy dose coloured by allocated regimen (pink: *Reg-DI*, purple: *Reg-C*) to show the different time schedule of the two arms. Patients - also within the same regimen - reported different values of standardized cumulative dose during time, depending on the delays and dose reductions required during chemotherapy due to toxicity. In particular, the lines form a tight bundle in the early phase of the treatment, but later they open up in a hand-fan shape because treatment adjustments are generally more frequent towards the end of the protocol.

The association between the longitudinal trajectories of time-varying covariates and the time-to-event outcome is discussed in the next section.
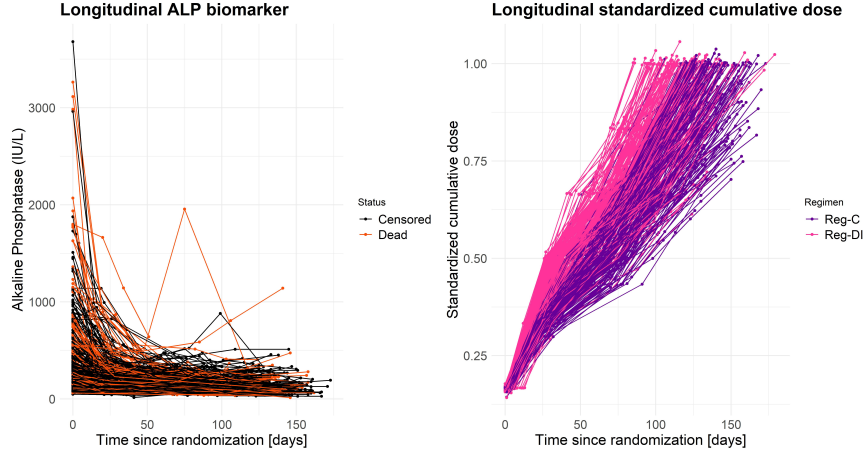
Figure 3: Time-varying covariates for each patient. Left panel: longitudinal values of ALP biomarker over time coloured by event status (black: *Censored*, red: *Dead*). Right panel: longitudinal values of standardized cumulative dose of chemotherapy coloured by allocated regimen (pink: *Reg-DI*, purple: *Reg-C*)

4 Results

In this section the methodology proposed in Section 2 is applied to the MRC BO06 osteosarcoma trial. First a joint modelling analysis for the internal longitudinal ALP biomarker and time-to-death is performed (Section 4.1). Then, both time-varying ALP biomarker and time-varying standardized cumulative dose of chemotherapy are modelled into a multivariate FLCRM (Section 4.2). Based on clinical information, each model is adjusted for three baseline covariates: age (< 15 or ≥ 15), gender and White Blood Count [$\times 10^9/L$] at randomization.

4.1 Joint modelling of longitudinal ALP biomarker and time-to-death

In the joint modelling analysis the choice of the longitudinal sub-process (1) for endogenous alkaline phosphatase was driven by the nature of the time-varying process. Since the longitudinal trajectories of the biomarker are skewed and non-linear, the logarithm of ALP was incorporated in the model and natural cubic splines for the fixed and random effects parts of the process were used. For patient i the longitudinal process was defined as:

$$\log\left[y_i(t)\right] = \eta_i(t) + \varepsilon_i(t) = (\beta_0 + b_{i0}) + \sum_{k=1}^{2}(\beta_k + b_{ik})B_n(t, \lambda_k) + \varepsilon_i(t) \tag{9}$$

where $\{B_n(t, \lambda_k) : k = 1, 2\}$ denotes the B-spline basis matrix for a natural cubic spline of time t with an internal knot placed at 50th percentile of the follow-up times, $\varepsilon_i(t) \sim N(0, \sigma_\varepsilon^2 \mathbf{I}_{n_i})$ is the unknown vector of random errors and $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$ is the vector of the patient-specific random effects, with $\mathbf{D}$ unstructured variance-covariance matrix.

For the event sub-process (2), the hazard was assumed to be dependent on the current value of the subject-specific linear predictor $\eta_i(t)$ at time t adjusted for age, gender and WBC at baseline as follows:

$$h_i(t|\mathcal{H}_i(t), \boldsymbol{\omega}_i) = h_0(t)\exp\left\{\gamma_1 \, WBC_i + \gamma_2 \, gender_i + \gamma_3 \, age_i + \alpha_1 \eta_i(t)\right\}. \tag{10}$$

9

	HR	2.5% CI	97.5% CI
WBC_i	1.1288	1.0300	1.2337
Gender (male)	1.8999	1.1376	3.2376
Age (≥ 15)	0.7197	0.4245	1.1538
$\eta_i(t)$	1.0807	1.0094	1.2235

Table 2: Hazard Ratios (HR) along with their 95% credibility intervals (CI) for the event process estimated with the joint modelling.

Hazard Ratios (HR) and 95% credibility intervals (CI) for the event process, under the joint modelling analysis (9)–(10), are shown in Table 2. Being a male corresponded to 1.9-times faster experience of the event and a unit increase in the value of WBC corresponded to a 12.9% increase in the risk of death. No age-related effect was observed. HR related to the subject-specific linear predictor $\eta_i(t)$ at time t was $1.0807 > 1$, indicating that a unit increase of $\eta_i(t)$ corresponds to a 8.1% increase in the risk of death. Therefore, having a higher current value of the subject-specific linear predictor at time t is associated to a lower survival probability.

Through joint models it was also possible to perform a dynamic monitoring of the patients according to their longitudinal values at different time points. Figure 4 panel [A] shows the longitudinal values $\log[y_i(t)]$ and $\log[y_j(t)]$ and the relative subject-specific linear predictors $\eta_i(t)$ and $\eta_j(t)$ over time, estimated through the longitudinal sub-model (9), for two specific patients i and j with the same risk factors (male patients aged ≥ 15, with WBC at randomization equal to 7.65×10^9/L) but with different longitudinal trajectories over time. Patient i presented lower values of the estimated subject-specific linear predictor $\eta_i(t)$ compared to the values $\eta_j(t)$ of patient j at the same time points, and levels for both patients decreased over time. Figure 4 panel [B] shows the survival estimated at different time points $\boldsymbol{t} = (t_1, t_2, t_3) = (21, 77, 119)$ days. The survival of each patient changes according to the evolution of the biomarker during follow up. Patient i with lower values of $\eta_i(t)$ shows good estimated survival with smaller confidence intervals over time compared to patient j with high value of the linear predictor. The survival of patient j improves as long as the internal ALP biomarker level decreases, suggesting that incorporate in the model the evolution of the marker provide more information than consider only the baseline value.

In the next section the association among internal time-varying ALP biomarker, external standardized cumulative dose of chemotherapy and survival by using FDA is investigated.

4.2 Multivariate functional linear Cox regression model

In this section the multivariate FLCRM, with time-varying ALP biomarker and time-varying standardized cumulative dose of chemotherapy as risk factors, is estimated. First, the longitudinal processes introduced in Section 3.2 are represented in the form of functions (Section 4.2.1). Then, through Functional Principal Component Analyses (FPCA) (Section 4.2.2) they are incorporated into a multivariate FLCRM (Section 4.2.3).

4.2.1 From longitudinal to functional representation

To convert the longitudinal ALP values $\mathbf{y}_i$ into the functions $\tilde{x}_i^{(ALP)}(t)$, FDA techniques, as discussed by Ramsay and Silverman,[21,22] were exploited. Measurements by cycles and not by days were used. In this way all the time-varying values were on the same temporal domain, i.e. $S_{ALP} = [1, 6]$ cycles. Then, B-spline basis functions Φ (2 or 3 basis of order 2 or 3, according to each patient i) were chosen as basis function system and the functional data objects were
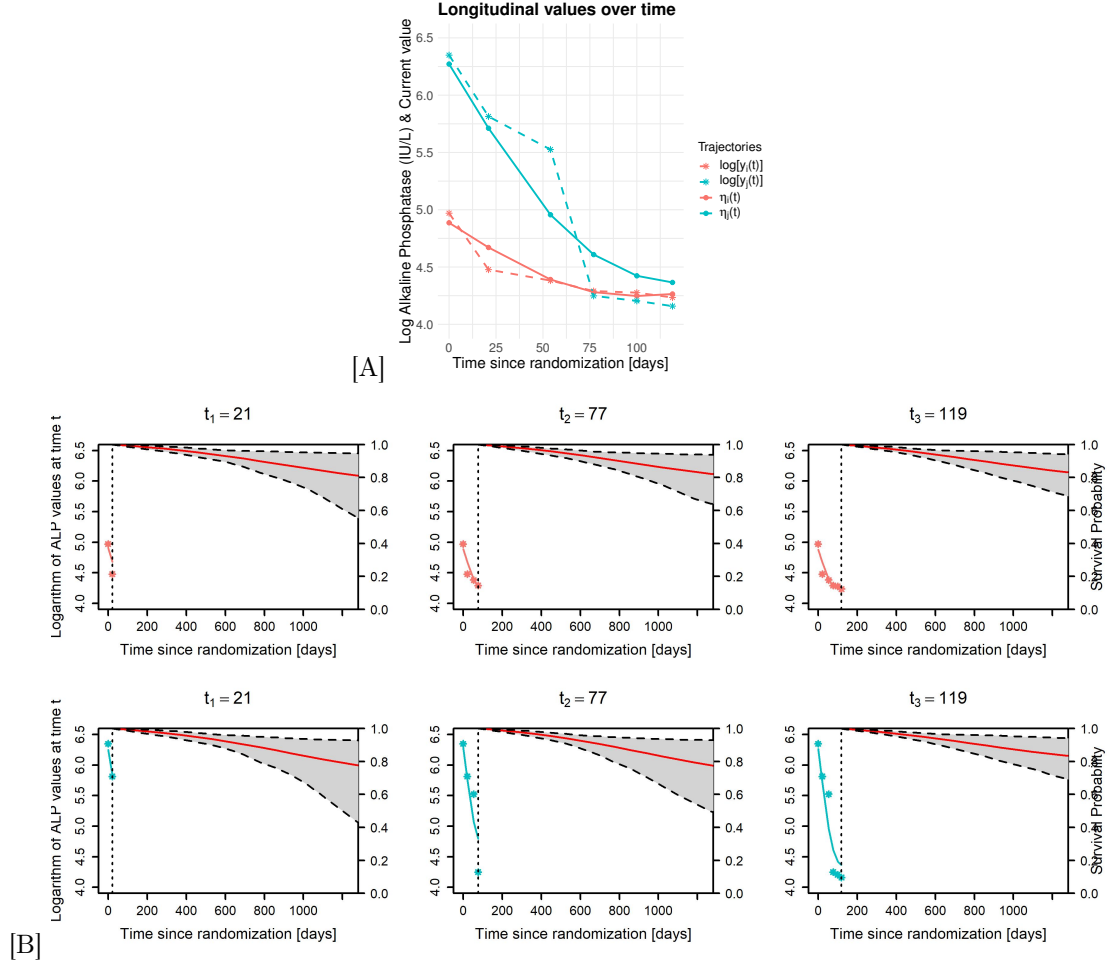
Figure 4: Panel [A]: Longitudinal values of $\log[y_i(t)]$ and $\log[y_j(t)]$ (dashed lines) and relative subject-specific linear predictors $\eta_i(t)$ and $\eta_j(t)$ (solid lines) for two different patients i (pink) and j (light-blue). Panel [B]: Survival probability plots at different time points for patient i (top panels) and j (bottom panels). Time $t_0 = 0$ corresponds to the time of randomization. *Left panels:* longitudinal process is considered from t_0 to $t_1 = 21$ days and survival is estimated starting from t_1. *Central panels:* longitudinal process is considered from t_0 to $t_2 = 77$ days and survival is estimated starting from t_2. *Right panels:* longitudinal process is considered from t_0 to $t_3 = 119$ days and survival is estimated starting from t_3.

11

expressed as linear combination of the basis functions. Finally, data were smoothed by regression analysis using the transformation $\tilde{g}(x) = \log \frac{x-L}{U-x}$ in order to constrain functions between a lower $L = 0\,IU/L$ and an upper $U = 4000\,IU/L$ bound. For each patient i the following functional ALP predictor was provided:

$$\tilde{x}_i^{(ALP)}(t) = \frac{4000 \cdot \exp\left[\boldsymbol{\phi}_i(t)^T \widehat{\mathbf{c}}_i\right]}{1 + \exp\left[\boldsymbol{\phi}_i(t)^T \widehat{\mathbf{c}}_i\right]} \tag{11}$$

where $\phi_i(t)$ and $\widehat{\mathbf{c}}_i$ are the vectors of B-spline basis functions at time t and estimated coefficients, respectively. A graphical representation of functional ALP biomarker curves $\tilde{x}_i^{(ALP)}(t)$ is reported in left panel of Figure 5. Each curve represents the functional ALP predictor of a patient coloured according to the event status.

To convert the longitudinal values of standardized cumulative chemotherapy dose $\boldsymbol{\delta}_i$ into the functional form $\tilde{x}_i^{(\delta)}(t)$, measurements in days were considered since different duration in treatment is a key-point of the chemotherapy protocol. In particular, the interval $S_\delta = [0, 180]$ days was selected, since all the patients completed the therapy within 180 days from randomization. Then, B-spline basis functions Φ (5 basis of order 5) were chosen as our basis function system, while functional data objects was defined by a monotone form, as discussed in Ramsay and Silverman.[21,22] Finally, data were smoothed by penalized regression analysis using the transformation $\tilde{g}(x)$ in order to constrain functions between a lower $L = 0$ and an upper $U = 1.1$ bound. For each patient i a functional predictor of standardized cumulative dose of chemotherapy was obtained:

$$\tilde{x}_i^{(\delta)}(t) = \frac{1.1 \cdot \exp\left(\widehat{\beta}_{0i} + \widehat{\beta}_{1i} \int_0^t \exp\left[\boldsymbol{\phi}(u)^T \widehat{\mathbf{c}}_i\right] du\right)}{1 + \exp\left(\widehat{\beta}_{0i} + \widehat{\beta}_{1i} \int_0^t \exp\left[\boldsymbol{\phi}(u)^T \widehat{\mathbf{c}}_i\right] du\right)} \tag{12}$$

where $\phi(t)$ and $(\widehat{\beta}_{0i}, \widehat{\beta}_{1i}, \widehat{\mathbf{c}}_i)$ are the vectors of B-spline basis functions at time t and estimated coefficients respectively. A graphical representation of functional standardized cumulative dose curves $\tilde{x}_i^{(\delta)}(t)$ is shown on the right panel of Figure 5. Each line represents the functional predictor of standardized cumulative dose for patient i coloured according to the allocated regimen.

Figure 5 shows that a functional representation of the time-varying processes highlights trends and variations in the shape of processes over time.

4.2.2 Functional principal component analysis for time-varying ALP biomarker and chemotherapy dose

The functional predictors $\tilde{x}_i^{(ALP)}(t)$ and $\tilde{x}_i^{(\delta)}(t)$ for patient i provided in (11) and (12) respectively, were summarised into a finite set of covariates by applying Functional Principal Component Analyses (FPCAs). In both cases, two principal components were enough to consider at least 95% of the explained variance.

Results of FPCA on functional ALP predictors are provided in Figure 6. Panel [A] shows the first two Principal Components $\xi_k^{(ALP)}$ and the average ALP curve $\mu^{(ALP)}(t) \pm \sqrt{\lambda_k^{(ALP)}} \cdot \xi_k^{(ALP)}$. Panel [B] reports the FPC scores plot $(f_{i1}^{(ALP)}, f_{i2}^{(ALP)})$ with relative boxplots. Each point represents a patient coloured by status. The first component $\xi_1^{(ALP)}$ explained 84.7% of the variability and a positive score reflected higher values of ALP during time compared to the mean (top panels in [A]). In particular, the higher the score, the higher the ALP levels during the first period of the treatment. The second component $\xi_2^{(ALP)}$ explained 13% of the variability and positive scores reflected "flat" curves, whereas negative score reflected curves with high slope, especially in the first period (bottom panels in [A]).
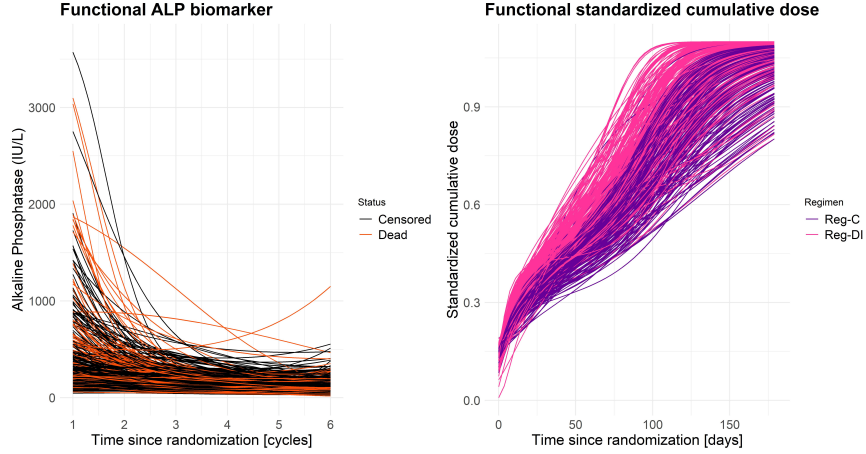
Figure 5: Left panel: functional representations of ALP biomarker over cycles coloured by status (black: *Censored*, red: *Dead*). Right panel: functional representations of standardized cumulative dose of chemotherapy over time coloured by allocated regimen (pink: *Reg-DI*, purple: *Reg-C*). Each line is the graphical representation of the functional predictor of a patient.

Results of FPCA on functional standardized cumulative dose $\tilde{x}_i^{(\delta)}(t)$ are shown in Figure 7. Panel [A] shows the first two Principal Components $\xi_k^{(\delta)}$ and the average ALP curve $\mu^{(\delta)}(t)$ $\pm 3\sqrt{\lambda_k^{(\delta)}} \cdot \xi_k^{(\delta)}$. Panel [B] reports the FPC scores plot $(f_{i1}^{(\delta)}, f_{i2}^{(\delta)})$ with relative boxplots. Each point corresponds to a patient. Different colours represent the two regimens. The first component $\xi_1^{(\delta)}$ explained 87.1% of the variability and a positive score reflected a shorter duration of the treatment (top panels in [A]). The second component $\xi_2^{(\delta)}$ explained 9.9% of the variability and a positive score indicated a faster growth in the chemotherapy assumption in the first period compared to the second one, with respect to the mean (bottom panels in [A]). Every two patients reported different values of FPC scores, reflecting delays or dose reductions during chemotherapy. This representation is able to represent different treatment dynamics, also among patients allocated the same regimen.

Boxplots in Figures 6 and 7 shows that scores related to FPCA on ALP biomarker had larger ranges of values respect to the ones related to functional chemotherapy dose, reflecting differences in ranges of the functional predictors. FPC scores $(f_{i1}^{(ALP)}, f_{i2}^{(ALP)}, f_{i1}^{(\delta)}, f_{i2}^{(\delta)})$ are used in Section 4.2.3 for the multivariate FLCRM analysis.

4.2.3 Multivariate functional linear Cox regression model

For functional representations and FPCA the *pre-defined period* $[T_0; T_0^*]$ composed by the first 180 days after randomization was considered. To study the effect of risk factors on survival from T_0^* in Figure 1, the following model was estimated:

$$
\begin{aligned}
h_i\left(t|\boldsymbol{\omega}_i, \tilde{x}_i^{(\delta)}(t), \tilde{x}_i^{(ALP)}(t)\right) = h_0(t)\exp\Big\{ & \gamma_1\,WBC_i + \gamma_2\,gender_i + \gamma_3\,age_i + \\
& \alpha_1^{(\delta)}f_{i1}^{(\delta)} + \alpha_2^{(\delta)}f_{i2}^{(\delta)} + \\
& \alpha_1^{(ALP)}f_{i1}^{(ALP)} + \alpha_2^{(ALP)}f_{i2}^{(ALP)} \Big\}
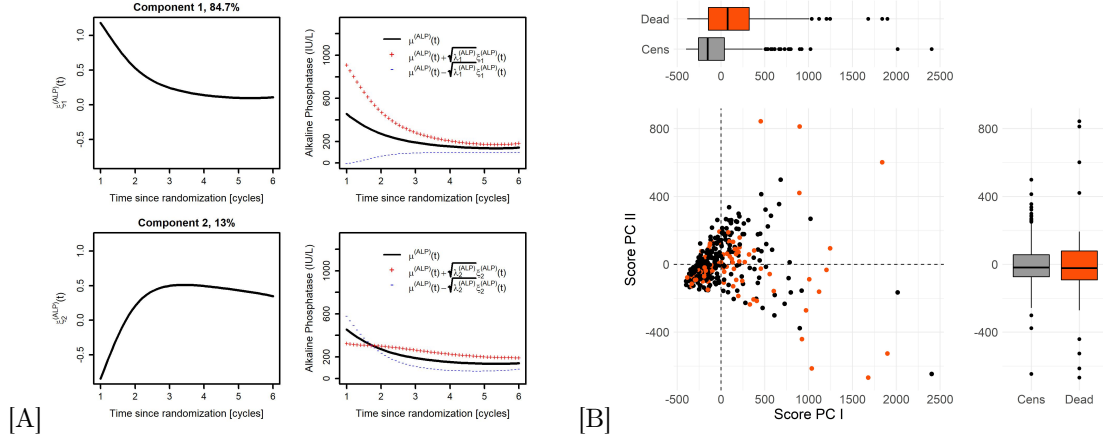\end{aligned} \tag{13}
$$

13

Figure 6: FPCA for functional Alkaline Phosphatase $\tilde{x}_i^{(ALP)}(t)$. [A] Left panel: Principal Components $\xi_k^{(ALP)}$ ($k = 1, 2$). Right panel: average ALP curve $\mu^{(ALP)}(t) \pm \sqrt{\lambda_k^{(ALP)}} \cdot \xi_k^{(ALP)}$.
[B] Functional PC scores plot $\left(f_{i1}^{(ALP)}, f_{i2}^{(ALP)} \right)$ with boxplots (black: *Censored*, red: *Dead*).
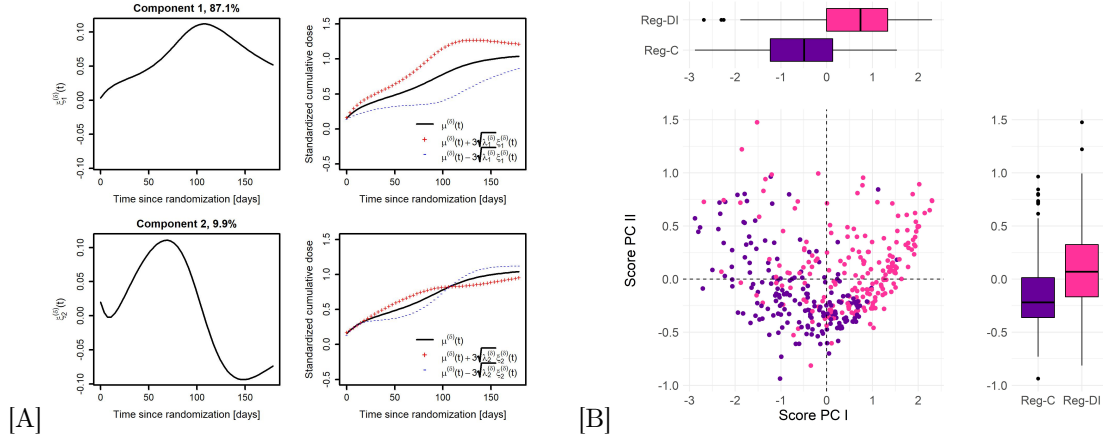

Figure 7: FPCA for functional standardized cumulative dose $\tilde{x}_i^{(\delta)}(t)$. [A] Left panel: Principal Components $\xi_k^{(\delta)}$ ($k = 1, 2$). Right panel: average standardized cumulative dose $\mu^{(\delta)}(t) \pm 3 \sqrt{\lambda_k^{(\delta)}} \cdot \xi_k^{(\delta)}$.
[B] Functional PC scores plot $\left(f_{i1}^{(\delta)}, f_{i2}^{(\delta)} \right)$ with boxplots (pink: *Reg-DI*, purple: *Reg-C*).

14

	HR	95% CI
WBC	1.1135	[1.0248; 1.2099]
Gender (male)	1.6900	[1.0258; 2.7842]
Age (≥ 15)	0.7669	[0.4673; 1.2588]
$f_1^{(\delta)}$	0.9335	[0.7586; 1.1486]
$f_2^{(\delta)}$	0.7894	[0.4358; 1.4299]
$f_1^{(ALP)}$	1.0011	[1.0006; 1.0015]
$f_2^{(ALP)}$	0.9998	[0.9986; 1.0011]

Table 3: Estimated hazard ratios (HR) along with 95% confidence intervals (CI) from a multivariate functional linear Cox regression model.

where $\boldsymbol{\omega}_i = (WBC_i, gender_i, age_i)$ is the vector of baseline covariates, $\tilde{x}_i^{(ALP)}(t)$ and $\tilde{x}_i^{(\delta)}(t)$ are the functional predictors of ALP and standardized cumulative dose respectively, with relative FPC scores $f_{ik}^{(m)}$ ($k = 1, 2; m \in \{ALP, \delta\}$).

In Table 3 hazard ratios along with their 95% confidence interval are shown. Gender, level of WBC at randomization and the score related to the first PC of alkaline phosphatase $f_{i1}^{(ALP)}$ were associate to survival. Being a male corresponded to 1.7-times faster experience of the event. A 1-unit increase in the value of WBC was associated to a 11.35% increase in the risk of death. Patients with high ALP levels had poor survival, as also resulted from the joint modelling analysis. FPC scores related to functional chemotherapy dose showed no effects on survival. Although all longitudinal information concerning a patient are incorporated in the model, dose-intense profiles did not show an improvement on survival.

Estimated survival probabilities based on the multivariate FLCRM are shown in Figure 8. High values of WBC corresponded to poor survival (top-left panel). Male patients had a low survival (top-right panel). The effect of functional ALP biomarker $f_{i1}^{(ALP)}$ suggested that patient i with high ALP levels over time had high risk of death (bottom-left panel). The score $f_{i1}^{(\delta)}$ related to the first PC of functional chemotherapy indicated that there was no improvement on survival due to dose-intense profiles (bottom-right panel).

5 Discussion

To study the association between time-varying processes and time-to-event data is a challenging problem in clinical research and the development of models and methods able to properly deal with dynamic time-varying covariates is of statistical interest and of clinical relevance. Research into joint modelling of longitudinal and time-to-event data or functional modelling have received considerable attention in recent years. In this work, a new approach to investigate the dynamics of time-varying processes over time and to include additional information that may be related to the survival into the time-to-event model was presented. Data from the MRC BO06/EORTC 80931 randomized clinical trial for osteosarcoma treatment were analysed. Since the 5-year overall survival for osteosarcoma patients remained unchanged in the past 40 years, it is extremely important to develop a new methodology to evaluate the prognosis and to define the management approach. In cancer studies, biomarkers or chemotherapy regimens are usually included in survival models as fixed baseline covariates. This approach does not take into account the evolution of the treatment process for each patient, discarding valuable information. Therefore, ALP biomarker values and chemotherapy dose during treatment as time-varying covariates were incorporated into time-to-event models using two different methodologies. A joint modelling technique
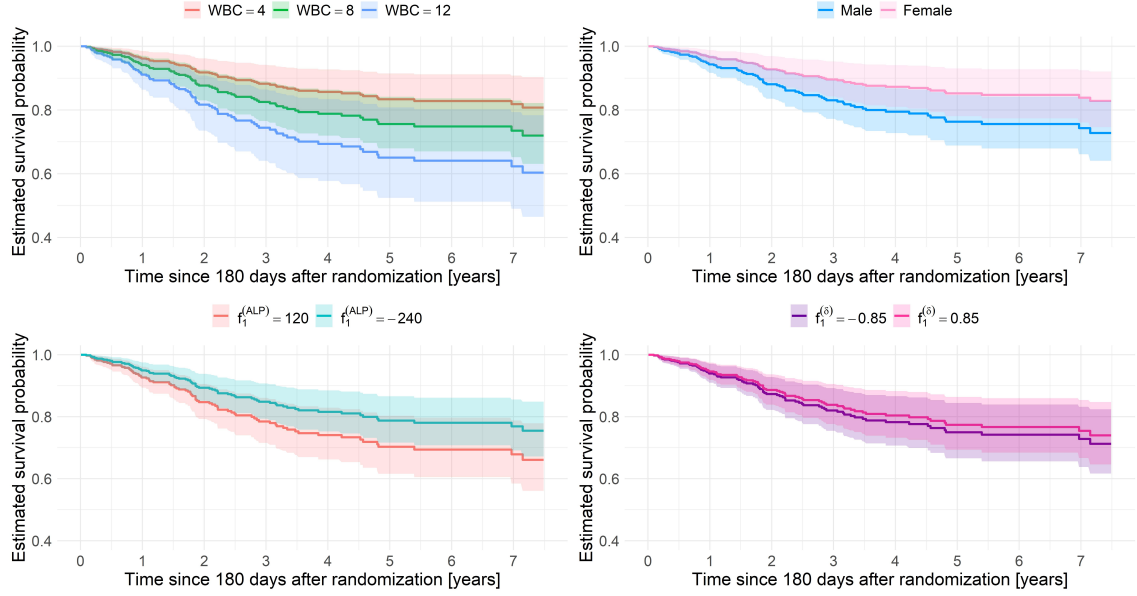
Figure 8: Estimated survival probability based on the multivariate functional linear Cox regression model (13). Time $t_0 = 0$ corresponds to T_0^* in Figure 1. Top-left panel: patients with different values of WBC $[\times 10^9/\text{L}]$ at randomization (red: $WBC = 4$; green: $WBC = 8$; blue: $WBC = 12$). Top-right panel: *female* (pink) vs *male* (blue) patients. Bottom-left panel: patients with different values of the first PC score for functional ALP biomarker (sea-green: $f_1^{(ALP)} = -240$; red: $f_1^{(ALP)} = 120$). Bottom-right panel: patients with different values of the first PC score for functional chemotherapy (purple: $f_1^{(\delta)} = -0.85$; pink: $f_1^{(\delta)} = 0.85$). When not specified, the other risk factors are fixed to the most frequent class for categorical covariates, i.e. *males* aged ≥ 15, and to the median value for continuous ones, i.e. $WBC = 7.65 \times 10^9/L$ at randomization, $f_1^{(\delta)} = 0.09$, $f_2^{(\delta)} = -0.08$, $f_1^{(ALP)} = -127$ and $f_2^{(ALP)} = -19$.

16

was first used to investigate how patients' survival were influenced by internal ALP longitudinal values. Then, FDA techniques were exploited to represent both internal time-varying biomarker and external chemotherapy dose in terms of functions, able to reflect trends and variations in the evolution of the processes over time. To study the effects of functional risk factors on survival, FLCRM was extended to the case of multivariate functional predictors and finally estimated.

Both approaches led to similar results. Suggesting that osteosarcoma patients with high ALP levels during time have poor overall survival, important information about the longitudinal and functional behaviours of the generating processes that underpin the data were taken into account, representing interpretative and forecasting tools in osteosarcoma research. On one hand, the joint modelling approach allowed to directly monitoring the effects of ALP levels on patient's survival. This real-time monitoring and profiling of patients, crucial for subject-specific predictions and personalized medicine, could allow to tailor therapeutic interventions in order to prevent disease progression. On the other hand, FDA methods allowed to extract additional information contained in the functions, representing an effective exploratory and modelling technique for highlighting trends and variations in the evolution of the processes over time. Exploiting the individual progression of ALP levels allowed to investigate their effects on survival instead of using an approach that consider only baseline information. Similarly, although dose-intense profiles were not associated with a better survival, functional chemotherapy representation was able to capture individual realisations of the intended treatment, which depends on delays and dose reductions reported throughout the course of treatment, detecting differences between patients randomised to the same regimen.

Dose-intense profiles were not associated with survival, even if functional chemotherapy representations were able to capture individual realisations of the intended treatment, suggesting that considering only the assumed dose as treatment proxy is not enough. Chemotherapy presents some particular aspects, such as latent accumulation of toxicity, which must be taken into account.[39] The use of FLCRMs could lead to a loss of information due to cohort selection. Indeed, it is necessary that patients survived for a period at least equal to the length of the *pre-defined period* used to compute the functional predictors. This could imply a loss of information if many patients are excluded. This issue can be overcame using time-varying covariates into an appropriate survival framework, like the joint modelling approach. However, a joint modelling approach is not appropriate to study the evolution of exogenous time-varying covariates.

The strength of our method is the ability to detect differences between patients with different biomarker evolution and treatment response, even when randomised to the same regimen. This work opens doors to many further developments, both in the field of statistical methods and in cancer research. From a clinical point of view, it will be necessary to simultaneously consider chemotherapy treatment modifications and developments of toxicities or adverse events. This aspect need to be taken into account into the representation of the dynamic evolution of these processes. To model them simultaneously is not a trivial task. It could also be interesting to analyse different biomarkers, in order to detect their prognostic role. These measurements could be of clinical interest since they represent dynamic patterns that could reflect patient's disease progression. Studying factors that could be related to chemotherapy treatment might provide interesting analyses and strong external validity.

The complexity of the phenomenon asks for the developments of new methodologies, able deal with the complicated aspects of chemotherapy treatment. This study shows that working in this direction is a difficult but profitable approach, which could lead to new improvements for subject-specific predictions and personalised treatment.
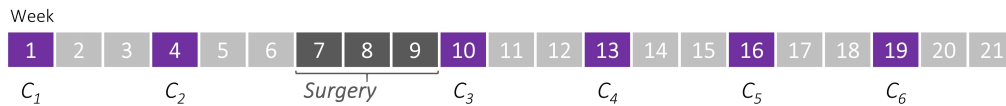
Acknowledgements

References

[1] Smeland S, Bielack S, Whelan J et al. Survival and prognosis with osteosarcoma: outcomes in more than 2000 patients in the EURAMOS-1 (European and American Osteosarcoma Study) cohort. *Eur J Cancer* 2019; 109: 36–50.

[2] Ritter J and Bielack S. Osteosarcoma. *Ann Oncol* 2010; 21(suppl 7): vii320–vii325.

[3] Lewis I, Nooij MA, Whelan J et al. Improvement in Histologic Response But Not Survival in Osteosarcoma Patients Treated With Intensified Chemotherapy: A Randomized Phase III Trial of the European Osteosarcoma Intergroup. *JNCI: J Natl Cancer Inst* 2007; 99(2): 112–128.

[4] Austin PC, Latouche A and Fine JP. A review of the use of time-varying covariates in the fine-gray subdistribution hazard competing risk regression model. *Stat Med* 2020; 39(2): 103–113.

[5] Kim S, Shin KH, Moon SH et al. Reassessment of alkaline phosphatase as serum tumor marker with high specificity in osteosarcoma. *Cancer Med* 2017; 6(6): 1311–1322.

[6] Zingler C and Schuff-Werner P. Significance of alkaline phosphatase isoenzymes in bone. *Dtsch Med Wochenschr* 2001; 126: 352.

[7] Hao H, Chen L, Huang D et al. Meta-analysis of alkaline phosphatase and prognosis for osteosarcoma. *Eur J Cancer Care* 2017; 26(5): e12536.

[8] Ren H, Sun LL, Li HY et al. Prognostic significance of serum alkaline phosphatase level in osteosarcoma: A meta-analysis of published data. *Biomed Res Int* 2015; 2015(Article ID 160835).

[9] Lancia C, Anninga J, Sydes MR et al. A novel method to address the association between received dose intensity and survival outcome: benefits of approaching treatment intensification at a more individualised level in a trial of the European Osteosarcoma Intergroup. *Cancer Chemother Pharmacol* 2019; 83(5): 951–962.

[10] Lancia C, Anninga J, Sydes MR et al. Method to measure the mismatch between target and achieved received dose intensity of chemotherapy in cancer trials: a retrospective analysis of the MRC BO06 trial in osteosarcoma. *BMJ Open* 2019; 9(5).

[11] Faucett C and Thomas D. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Stat Med* 1996; 15(15): 1663–1685.

[12] Tsiatis A and Davidian M. Joint modelling of longitudinal and time-to-event data: an overview. *Stat Sin* 2004; 14(3): 809–834.

[13] Chi Y and Ibrahim J. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics* 2006; 62(2): 432–445.

[14] Williamson P, Kolamunnage-Dona R, Philipson P et al. Joint modelling of longitudinal and competing risks data. *Stat Med* 2008; 27(30): 6426–6438.

[15] Liu L and Huang X. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *J R Stat Soc Ser C Appl Stat* 2009; 58(1): 65–81.

[16] Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data with Applications in R.* Chapman & Hall/CRC, 2012.

[17] Gould L, Boye ME, Crowther MJ et al. Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. *Stat Med* 2015; 34(14): 2181–2195.

[18] Hickey G, Philipson P, Jorgensen A et al. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Med Res Methodol* 2016; 16(117).

[19] Hickey G, Philipson P, Jorgensen A et al. A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial. *J R Stat Soc Ser A Stat Soc* 2018; 181(4): 1105–1123.

[20] Ramsay J and Dalzell C. Some tools for Functional Data Analysis. *J R Stat Soc Ser B Stat Methodol* 1991; 53: 539–572.

[21] Ramsay J and Silverman BW. *Applied Functional Data Analysis: Methods and Case Studies.* Springer Series in Statistics, Springer New York, 2002.

[22] Ramsay J and Silverman BW. *Functional Data Analysis.* Springer Series in Statistics, Springer New York, 2005.

[23] Ferraty F and Vieu P. *Nonparametric Functional Data Analysis: Theory and Practice.* Springer Series in Statistics, Springer New York, 2006.

[24] Ullah S and Finch C. Applications of Functional Data Analysis: A systematic review. *BMC Medical Res Methodol* 2013; 13(43).

[25] Ieva F, Paganoni AM, Pigoli D et al. Multivariate functional clustering for the morphological analysis of electrocardiograph curves. *J R Stat Soc Ser C Appl Stat* 2013; 62(3): 401–418.

[26] Ieva F and Paganoni AM. Risk prediction for myocardial infarction via generalized functional regression models. *Stat Methods Med Res* 2016; 25(4): 1648–1660.

[27] Martino A, Ghiglietti A, Ieva F et al. A k-means procedure based on a mahalanobis type distance for clustering multivariate functional data. *SMA* 2019; 28(2): 301–322.

[28] Gellar JE, Colantuoni E, Needham DM et al. Cox regression models with functional covariates for survival data. *Stat Modelling* 2015; 15(3): 256–278.

[29] Lee E, Zhu H, Kong D et al. BFLCRM: A Bayesian functional linear Cox regression model for predicting time to conversion to Alzheimer's disease. *Ann App Stat* 2015; 9(4): 2153–2178.

[30] Qu S, Wang J and Wang X. Optimal estimation for the functional Cox model, 2016. *Ann Stat* 2016; 44(4): 1708–1738.

[31] Kong D, Ibrahim JG, Lee E et al. FLCRM: Functional linear Cox regression model. *Biometrics* 2018; 74(1): 109–117.

[32] Kalbfleisch JD and Prentice RL. *The Statistical Analysis of Failure Time Data.* Wiley Series in Probability and Statistics, John Wiley & Sons, 2011.

[33] Rizopoulos D. The R Package JMbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC. *J Stat Softw, Articles* 2016; 72(7): 1–46.

[34] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2018. URL `https://www.R-project.org/`.

[35] Ibrahim J, Chen M and Sinha D. *Bayesian Survival Analysis.* Springer-Verlag, New York, 2001.

[36] Brown E, Ibrahim J and DeGruttola V. A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics* 2005; 61: 64–73.

[37] Cox D. Regression models and life-tables. *J R Stat Soc* 1972; 34(2): 187–220.

[38] Schemper M and Smith T. A note on quantifying follow-up in studies of failure time. *Control Clin Trials* 1996; 17(4): 343–346.

[39] Lancia C, Spitoni C, Anninga C et al. Marginal structural models with dose-delay joint-exposure for assessing variations to chemotherapy intensity. *Stat Methods Med Res* 2019; 28(9): 2787–2801.

A   MRC BO06/EORTC 80931 RCT protocol

Data from the MRC BO06/EORTC 80931 Randomized Controlled Trial (RCT) for patients with non-metastatic high-grade osteosarcoma recruited between 1993 and 2002 were analysed.[3] The trial randomised patients between conventional treatment with doxorubicin (DOX) and cisplatin (CDDP) given every 3 weeks (*Reg-C*) versus a dose-intense regimen of the same two drugs given every 2 weeks (*Reg-DI*), supported by granulocyte colony-stimulating factor. Chemotherapy was administered for six cycles (a cycle is a period of either 2 or 3 weeks depending on the allocated regimen), before and after surgical removal of the primary osteosarcoma. In both arms, DOX (75 mg/m^2) plus CDDP (100 mg/m^2) were given over six cycles. Surgery to remove the primary tumour was scheduled at week 6 after starting treatment in both arms, that is, after 2 cycles ($2 \times$ [DOX+CDDP]) in regimen-C and after 3 cycles ($3 \times$ [DOX+CDDP]) in regimen-DI. Postoperative chemotherapy was intended to resume 3 weeks after surgery in both arms. Figure 9 shows the trial design. Laboratory tests, such as ALP test, were performed before each cycle in order to monitor patient's health status and the development of toxicities or adverse events. Delays or chemotherapy dose reductions during treatment were possible in case of toxicity. Additional details can be found in the primary analysis of the trial.[3]

Figure 9: Patients are randomized at baseline to one of the two regimens, with the same anticipated cumulative dose but different duration.

MOX Technical Reports, last issues

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

26/2020 Zonca, S.; Antonietti, P.F.; Vergara, C.
A Polygonal Discontinuous Galerkin formulation for contact mechanics in fluid-structure interaction problems

Regazzoni, F.; Dedè, L.; Quarteroni, A.
Biophysically detailed mathematical models of multiscale cardiac active mechanics

25/2020 Calvetti, D.; Cosmo, A.; Perotto, S.; Somersalo, E.
Bayesian mesh adaptation for estimating distributed parameters

24/2020 Formaggia, L.; Scotti, A.; Fumagalli, A.
Numerical Methods for Flow in Fractured Porous Media

23/2020 Spreafico, M.; Ieva, F.
Functional modelling of recurrent events on time-to-event processes

22/2020 Zeni, G.; Fontana, M.; Vantini, F.
Conformal Prediction: a Unified Review of Theory and New Challenges

21/2020 Benacchio, T.; Bonaventura, L.; Altenbernd, M.; Cantwell, C.D.; Düben, P.D.; Gillard, M.; Gir
Resilience and fault-tolerance in high-performance computing for numerical weather and climate prediction

20/2020 Almi, S.; Belz, S.; Micheletti, S.; Perotto, S.
A DIMENSION-REDUCTION MODEL FOR BRITTLE FRACTURES ON THIN SHELLS WITH MESH ADAPTIVITY

19/2020 Stella, S.; Vergara, C.; Maines, M.; Catanzariti, D.; Africa, P.; Demattè, C.; Centonze, M.; Nob
Integration of maps of activation times in computational cardiac electrophysiology

16/2020 Paolucci, R.; Mazzieri, I.; Piunno, G.; Smerzini, C.; Vanini, M.; Ozcebe, A.G.
Earthquake ground motion modelling of induced seismicity in the Groningen gas field