TESTING FOR THE FUNCTIONAL FORM OF A CONTINUOUS COVARIATE IN THE SHARED-PARAMETER JOINT MODEL

Xavier Piulachs, Anouar El Ghouch, Ingrid Van Keilegom

REPRINT | 2025 / 26

UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication

Testing for the functional form of a continuous covariate in the shared-parameter joint model

Xavier Piulachs*[1] | Anouar El Ghouch[2] | Ingrid Van Keilegom[2,3]

[1]Department of Statistics and Operations Research, Polytechnic University of Catalonia, Campus Terrassa, Spain
[2]Institute of Statistics, Biostatistics and Actuarial Sciences, UCLouvain, Louvain-la-Neuve, Belgium
[3]Research Centre for Operations Research and Statistics, KU Leuven, Leuven, Belgium

Correspondence
*Xavier Piulachs, Department of Statistics and Operations Research, Polytechnic University of Catalonia, Campus Terrassa, Building TR5, C. Colom 11, 08222 Terrassa, Spain. Email: xavier.piulachs@upc.edu

Summary

Shared-parameter joint modeling is a useful technique for properly associating longitudinal and time-to-event data. When the interest lies in the survival outcome, the conditional logarithm of the hazard function for an event is conventionally presumed to be linearly related over time to, among other terms, a set of fixed covariates. However, this hypothesis is highly restrictive and may yield misleading results. Our objective here is to easily check such a modeling assumption for any continuous fixed covariate. For this purpose, we examine the appropriateness of a nonparametric test criterion based on a penalty-modified version of the Akaike information criterion. An extensive numerical study is conducted to check the validity of the test within the joint modeling framework, while simultaneously determining the degree to which the function embedding a continuous covariate of interest deviates from linearity. Furthermore, once a deviation from linearity is detected, the improvement in the model's predictive performance is examined. The usefulness of the testing procedure is illustrated using a clinical trial conducted on HIV-infected subjects. Specifically, our example focuses on properly accounting for the effect of nadir CD4 cell count within a predictive joint model for the time to immune recovery.

KEYWORDS:
Akaike information criterion, order selection test, nonlinear covariate, joint model

1 | INTRODUCTION

To properly check the shape that underlies any covariate effect in a hazard regression model, nonparametric testing procedures are appealing from a practical standpoint. Such procedures postulate a flexible alternative class of models, theoretically characterized by an infinite number of parameters, and are designed to be consistent in the sense of their power against essentially any defined alternative that deviates from linearity.[1,2,3,4,5,6,7] The current article addresses the nonparametric checking of the functional shape for any continuous covariate within the scope of shared-parameter (SP) joint models for longitudinal and time-to-event data.[8,9,10,11] Joint modeling is required when the subject-specific longitudinal trajectory of a biomarker is strongly predictive of the time at which a predefined endpoint occurs. The time-dependent covariate defined by the longitudinal response, typically measured with error, is incorporated into a relative risk model through shared random effects between the longitudinal and time-to-event processes. While understanding the association between both processes is undoubtedly one of the more frequently analyzed features in joint modeling, other technical aspects deserve further investigation as well. Particularly, the conventional formulation for a relative risk model specifies that the conditional log-hazard for an event is linearly related to the prognostic covariates. This hypothesis, however, does not necessarily hold, and alternative relationships need to be tested. In practice,

though, it may be challenging to infer the appropiate functional form, say $f(\cdot)$, underlying a given covariate. A useful approach consists of testing the preconceived linear form against an alternative nonparametric hypothesis.[12,13]

Notably related to nonparametric smoothing procedures are the order selection tests, introduced by Eubank[14] to examine the adequacy of linear regression models. Here, the alternative hypothesis poses a, possibly orthogonal, series expansion of plausible basis functions. The orthogonality property, though not always absolutely necessary, is important from the standpoint of using the existing theory on distribution of the order selection test. Of course, another key motivation for using an orthogonal series expansion lies in guaranteeing that different basis functions included to approximate the function $f(\cdot)$ are not redundant, that is, all of them extend the null model by supplying independent information about this unknown function. The specific number of additional terms in the series expansion of $f(\cdot)$, hereinafter referred to as the order r of the test, could be subjectively selected from a graphical assessment, but a data-driven rule has proven to be more valuable. In a likelihood context, the order of the functional estimator is typically selected by using Akaike's information criterion (AIC),[15] under which the null hypothesis is rejected if and only if the data-driven integer r is greater than zero. Nevertheless, this popular selection criterion yields an asymptotic level quite far from the commonly accepted Type I error probability. To remedy this, we build on the proposal suggested by Aerts et al[16], which hinges upon a penalty-modified Akaike information criterion (MAIC). Hence, as an alternative to the standard formulation of SP joint models, a nonlinear flexible model can be considered for $f(\cdot)$ when relating the conditional log-hazard to a particular survival covariate. There are at least two works in survival analysis closely related to our research topic: Nysen, Aerts, and Faes[17] presented a lack-of-fit test for different types of random censorship, while Baayen, Hougaard, and Pipper[18] discussed a multiple testing procedure for analyzing dose-response studies. Nevertheless, to our knowledge, the performance of the MAIC-based test has not yet been analyzed in the context of joint models.

The use of the MAIC testing procedure is illustrated using a longitudinal observational study on a cohort of adults infected with human immunodeficiency virus (HIV) in South Africa.[19] After 2004, individuals started antiretroviral therapy (ART), established as the time zero, and key clinical information was monitored biannually after that for a period lasting up to five years. This monitoring included subject-specific biannual measurements of CD4 cell count (cells/μL), serving as a surrogate biomarker to describe the disease progression. Taking advantage of such longitudinal information, we focus on gaining insight into the well-recognized prognostic role played by the nadir CD4, defined as the lowest CD4 measurement recorded for an individual. To this end, the original data is reorganized for joint modeling techniques to be applied. Both the nadir CD4 and the time-dependent CD4 responses are employed to assess the individual log-hazard for immune recovery, identified when any longitudinal response satisfies CD4 $\geq$ 500 cells/μL. Thus, we are here concerned with a special case in which modeling the "risk" for a beneficial health event is of interest. For methodological reasons, we retain exclusively the CD4 measurements collected from $t = 0$ up to the earliest observed time at which the threshold of 500 cells/μL is surpassed, with the subsequent subject-specific CD4 observations then being discarded. Moreover, we excluded the following individuals: subjects with less than three measurements, subjects who had achieved immune recovery from the beginning, and subjects without the initial visit (and who are therefore subject to left truncation). Also, we removed subjects with more than a one-year gap of information before reaching a measurement showing immune recovery.

There are $n = 628$ subjects (68.6% females) ultimately included in our study, hereafter referred to as the HIV data. Let T_i^*, $i = 1, \ldots, n$, denote a particular event outcome, providing the point at which a subject-specific trajectory surpasses the limit of 500 cells/μL. To conceptualize the two different censoring patterns in the HIV data, T_i^* can be assumed to lie between a lower bound T_{Li}, at which the last available measurement under 500 cells/μL is collected, and an upper bound T_{Ui}, at which the earliest available value exceeding this limit occurs. Among all subjects, the observed trajectory for 425 (67.7%) of them did not surpass the prespecified threshold, T_i^* being right censored at $T_{Li} < T_i^*$, with $T_{Ui} \longrightarrow \infty$. The remaining 203 (32.3%) did, but their exact threshold-crossing time T_i^* is unknown because measurements were only taken every 6 months. For these subjects, only $T_i^* \in (T_{Li}, T_{Ui})$ can be guaranteed. In principle, this instance would lead to interval-censored event outcomes, where T_i^* could be imputed. In our case, however, only 11.3% of the 203 intervals are longer than 0.5 years, so hardly any overlap of survival information occurs. Our interval-censoring scheme belongs to the special case of severely grouped survival data,[20] and any imputation procedure performed on T_i^* becomes, in practical terms, totally random. We opted for setting $T_i^* = (T_{Li}+T_{Ui})/2$, this leading to a total median follow-up of 2.0 years. Further explanations on the HIV data are given in Section A1 of Appendix A, graphically illustrating the ordered censoring intervals (Figure S1). As regards the subject-specific evolution of longitudinal biomarker measurements, the left panel of Figure 1 depicts the CD4 longitudinal profiles observed for a set of subjects via biannual serological examination. A wide variability in subject CD4 responses is observed after initiating ART, partially due to possible adverse reactions.[21] Nonetheless, the average rate of change of the observed trajectories is likely to be greater for those HIV-infected subjects whose immune function is restored. Additionally, as our main interest, we examine the role played

by the nadir CD4 as an explanatory covariate for immune recovery. The right panel of Figure 1 shows the association between the Cox-based estimate of the nadir CD4, suitably rescaled to the unit interval [0, 1], and the log-hazard for immune recovery. In general, the relative risk shows a monotone pattern that increases in tandem with the covariate, which is logical since a higher nadir CD4 is typically correlated with a better health status and a higher propensity to achieve immune recovery. Nevertheless, a decreasing trend is observed near the lower boundary of [0, 1], in sharp contrast to the generally presumed linear effect on the conditional log-hazard function. Section A2 from Appendix A provides a list of the variables of the HIV data (Table S1), along with brief descriptive statistics (Tables S2 and S3). The results show that both subjects who achieve immune reconstitution and females exhibit markedly higher median nadir and overall CD4 values.

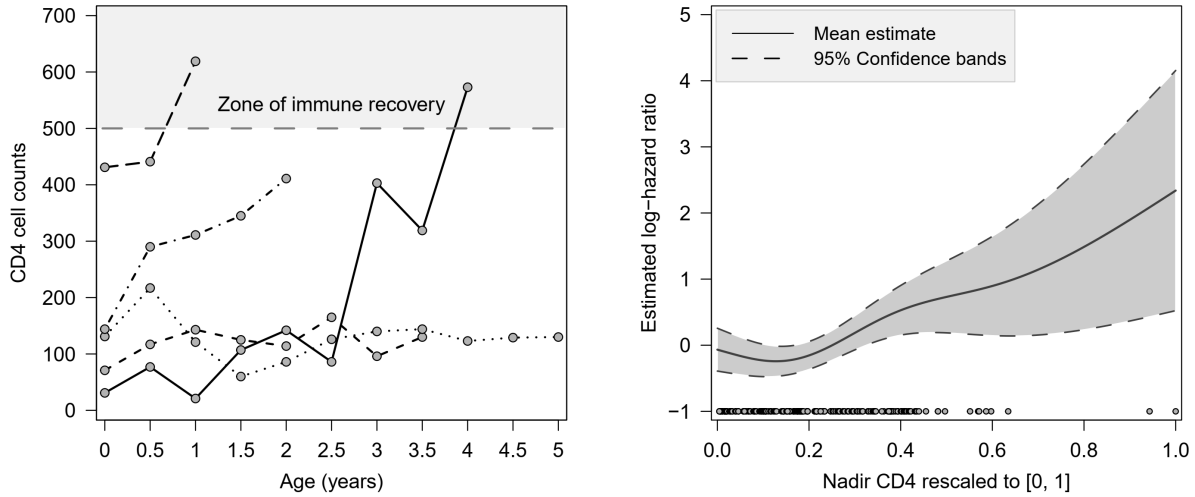

FIGURE 1 Left panel: Observed CD4 cell count measurements versus the follow-up time for five randomly chosen subjects. Right panel: Estimated log-hazard ratio as a function of the rescaled nadir CD4 cell count under a Cox approach. Points at the bottom of the plot indicate uncensored observations.

The rest of the article is organized as follows. Section 2 briefly introduces the mathematical formulation for the standard SP joint model when accounting for a single, nonlinear continuous covariate. Section 3 outlines how the lack-of-fit testing method proposed by Aerts et al [16] can be applied to suggest a proper functional shape for including a time-fixed covariate within the hazard function. Section 4 describes the parameter estimation of the SP joint model under a full likelihood paradigm. Section 5 provides a simulation study for examining the performance of the MAIC-based test within the scope of joint modeling techniques. In Section 6, the real example of HIV data is used to illustrate a brief application of the testing procedure within the joint modeling framework. Lastly, Section 7 provides a final discussion.

2 | COVARIATE ADJUSTMENT FOR SP JOINT MODELS

2.1 | Standard modeling approach

Many medical follow-up studies have increasingly targeted at examining the progression over time of an intermittently observed marker, which is possibly subject to measurement error, and the instantaneous hazard rate for an event of interest. Their underlying relationship is properly assessed at the subject-specific level by using an SP joint model for longitudinal and time-to-event data. A linear mixed-effects model (LMM) is typically used to describe the trajectory of the time-dependent marker, while its current effect on survival is simultaneously accounted for by a relative risk model. Hence, when conditioned on random effects, the natural logarithm of the hazard function is linearly related to an unknown function of time, and a set of survival covariates, among which one or more time-dependent biomarkers are included. Some fixed covariates, however, may not be linearly related to the conditional log-hazard function. This potential departure from the standard formulation is considered below when summarizing the SP joint model equations.

Let $\mathbf{y}_i = \{y_i(t_{ij}), i = 1, \ldots, n\}$ denote the observed longitudinal biomarker measurements for the ith subject, recorded at a sequential set of time points t_{ij}, $j = 1, \ldots, n_i$. The subject's responses are affected by biological measurement errors collected longitudinally in the vector $\boldsymbol{\varepsilon}_i = \{\varepsilon_i(t_{ij}), i = 1, \ldots, n\}$, where a particular component specifically denotes the deviation of an

observation $y_i(t)$ from the true unobserved biomarker response $\mu_i(t)$ at time t. For the remainder of this article, the longitudinal measurements are assumed to be normally distributed, $y_i(t) \sim \mathrm{N}\{\mu_i(t), \sigma^2_\varepsilon\}$, with σ^2_ε representing the constant within-subject variance. Accordingly, the LMM that summarizes the underlying evolution over time for each subject takes the form

$$\begin{aligned} &y_i(t) = \mu_i(t) + \varepsilon_i(t) = \mathbf{x}_i^\top(t)\boldsymbol{\beta} + \mathbf{z}_i^\top(t)\mathbf{b}_i + \varepsilon_i(t), \\ &\mathbb{E}\{y_i(t) \mid \mathbf{b}_i\} = \mu_i(t),\ \mathbb{V}\{y_i(t) \mid \mathbf{b}_i\} = \sigma^2_\varepsilon < \infty, \\ &\mathbf{b}_i \sim \mathrm{N}_q(\mathbf{0}, \mathbf{D}),\ \varepsilon_i(t) \sim \mathrm{N}(0, \sigma^2_\varepsilon). \end{aligned} \tag{1}$$

The terms $\mathbf{x}_i^\top(t)$ and $\mathbf{z}_i^\top(t)$ denote the ith subject's row vectors of the fixed- and random-effects design matrices, respectively, while $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\top$ is the unknown parameter vector for fixed effects and $\mathbf{b}_i = (b_{i0}, b_{i1}, \ldots, b_{iq})^\top$ is the vector for random effects. These random coefficients enable us to adequately characterize unobserved subject-specific features of a particular trajectory, and are typically assumed to follow a zero-mean multivariate normal distribution with an unspecified $(q+1)\times(q+1)$ covariance matrix, $\mathbf{D}$. Moreover, it is assumed that $\mathrm{cov}\{\mathbf{b}_i, \boldsymbol{\varepsilon}_i(t)\} = 0$ for all t, and $\mathrm{cov}\{\varepsilon_i(t), \varepsilon_i(t')\} = 0$ for $t \neq t'$. For the time-to-event outcome, we use T_i^* to denote the true event time for the ith subject, and C_i to represent the potentially related right-censoring time. These two random variables are assumed to be conditionally independent given the covariate history. In practice, we only observe $T_i = \min\{T_i^*, C_i\}$ and $\delta_i = I(T_i^* \leq C_i)$, where $I(\cdot)$ represents the indicator function that equals 1 if the event occurs, and 0 otherwise. For jointly modeling an unobservable longitudinal process and event history data, a shared set of time-independent random effects $\mathbf{b}_i$ is assumed. This way, the time-dependent effect of the error-free biomarker history $\mathcal{M}_i(t) = \{\mu_i(s),\ 0 \leq s \leq t\}$, is accounted for within the classical structure of a relative hazards regression model. To allow for generality, the conditional log-hazard function for the ith subject can be related to four distinct terms:

$$\log h_i\{t \mid \mathcal{M}_i(t), \mathbf{w}_i, \widetilde{\omega}_i\} = \log h_0(t) + \boldsymbol{\lambda}^\top \mathbf{g}\{\mu_i(t)\} + \boldsymbol{\gamma}_w^\top \mathbf{w}_i + f(\widetilde{\omega}_i), \quad t > 0. \tag{2}$$

On the one hand, the conditional log-hazard is linearly related to the three first terms, namely: (i) the logarithm of an arbitrary baseline hazard function $h_0(\cdot) \geq 0$, (ii) a vector-valued function $\mathbf{g} = \{g_1(\cdot), \ldots, g_\ell(\cdot)\}^\top$ of true biomarker response, containing g-parametric functions that are one-to-one associated to $\log h_i(\cdot)$ through the components from the vector of association parameters $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_\ell)^\top$, and (iii) a vector of covariates $\mathbf{w}_i$ with regression parameters vector $\boldsymbol{\gamma}_w$. On the other hand, as the main interest of this article, we allow for the existence of an additional term accounting for a single presumedly nonlinear covariate $\widetilde{\omega}_i$, related to the log-hazard through an arbitrary and unobservable continuously differentiable function $f(\cdot)$. This extends the traditional assumptions to account for covariate effects, and it is implicit that the proposed SP joint model reduces to its standard formulation when $f(\cdot)$ is linear. Though beyond the scope of our research, note that subsequent extensions including more than only one nonlinear covariate could also be considered.

Within the framework of SP joint models, the log-baseline hazard function can be modeled either parametrically or via more flexible specifications which may include a piecewise-constant function or B-spline basis functions. For reasons of computational burden, we model $h_0(\cdot)$ by means of a biparametric Weibull function of time, $h_0(t) = \nu\sigma_t t^{\sigma_t - 1}$, where $\nu > 0$ and $\sigma_t > 0$ represent the shape and scale parameters, respectively. Thereby, the underlying risk is modeled for all subjects as a power of time, so such a risk is likely to increase whenever $\sigma_t > 1$, while it diminishes if $0 < \sigma_t < 1$. For the particular case of $\sigma_t = 1$, function $h_0(\cdot)$ becomes constant. This automatically confers smoothness when modeling the baseline hazard for an event over time, while only requiring a couple of parameters to be estimated.

2.2 | The MAIC-based test procedure

Next, we introduce a convenient notation to describe the application of the nonparametric MAIC-based procedure in the context of SP joint models. We assume that covariate $\widetilde{\omega}_i$ is supported, without loss of generality, on $[0, 1]$, and has an arbitrary smooth function $f(\cdot)$, with at least one bounded derivative on $(0, 1)$. Our question of interest is testing any departure from linearity between $\log h_i(\cdot)$ and covariate $\widetilde{\omega}_i$ itself, that is, testing

$$\mathrm{H}_0 :\ f(\cdot) \in \mathcal{F}_0 \quad \text{versus} \quad \mathrm{H}_\mathrm{A} :\ f(\cdot) \notin \mathcal{F}_0, \tag{3}$$

where $\mathcal{F}_0 = \{\widetilde{\omega}_i \to c_1\widetilde{\omega}_i,\ c_1 \in \mathcal{C} \subseteq \mathbb{R}\}$ is the set of one-parameter linear functions. If H_0 is rejected in favor of H_A, then $f(\cdot)$ is not linear and any nonlinear shape could theoretically be considered for including $\widetilde{\omega}_i$. In practical terms, one may test any departure from H_0 by approximating $f(\cdot)$ with a nonparametric estimator, say $\varphi(\cdot)$, which linearly combines an undetermined number r of known basis functions $\{u_1(\cdot), \ldots, u_r(\cdot)\}$ of $\widetilde{\omega}_i$. This gives rise to a series expansion of $f(\cdot)$ written as

$$\varphi(\widetilde{w}_i \mid \mathbf{c}) = c_1 \widetilde{w}_i + \sum_{k=1}^{r} c_{1+k} u_k(\widetilde{w}_i),\ r \geq 1, \tag{4}$$

so testing H_0 reduces fundamentally to testing whether $c_2 = \ldots = c_{1+r} = 0$. The order r that determines the smoothness of $\varphi(\cdot)$ is to be estimated from the data according to an optimal model selection criterion. Ultimately, we seek to actually assess whether incorporating r additional basis components to the null function leads to a significantly enhanced approximation of $f(\cdot)$ by $\varphi(\cdot)$. The basis functions should be chosen so that $\varphi(\cdot) \longrightarrow f(\cdot)$ as $r \longrightarrow \infty$, provided that some regularity requirements are satisfied.[16]

Because (4) involves a series expansion, the selection of the candidate class of basis functions $\{u_r(\cdot),\ r \geq 1\}$ becomes crucial. Although any reasonable set of sufficiently smooth functions can in principle be admitted, it is advisable to start by using computationally simple functions. Particularly, series of low-order polynomials or trigonometric functions are likely to provide fair representations of an unknown functional form if it does not involve a heavily oscillating shape. These two examples, though not particularly original, entail a rich collection of basis functions entirely suited to validating the performance of the analyzed testing procedure. Further note that such representations are to be understood as purely auxiliary yet helpful models through which H_0 can be tested, and in no case does the purpose of this article consist of estimating the unknown $f(\cdot)$ embracing $\widetilde{w}_i$.

2.3 | Computation of the test statistic

Let $\mathcal{D}_n = \{(\mathbf{y}_i, T_i, \delta_i),\ i = 1, \ldots, n\}$ summarize the observed information for the n individuals. Also, let $\boldsymbol{\theta} = (\boldsymbol{\theta}_y^\top, \boldsymbol{\theta}_t^\top, \boldsymbol{\theta}_b^\top)^\top$ denote the unknown complete parameter vector of the SP joint model, where $\boldsymbol{\theta}_y = (\boldsymbol{\beta}, \sigma_\epsilon)^\top$ collects the parameters of the longitudinal submodel, $\boldsymbol{\theta}_t = (\boldsymbol{\gamma}_w, \nu, \sigma_t, \mathbf{c}, \boldsymbol{\lambda})^\top$ the parameters related to the time-to-event response for some number of components $r \in \{0, 1, \ldots\}$ retained by $\mathbf{c}$ in the estimation of $\varphi(\cdot)$, and $\boldsymbol{\theta}_b \equiv \mathbf{D}$ the shared variance components. Conditioned on the random effects, the rth order log-likelihood function of the full SP joint model is denoted by $L_r(\boldsymbol{\theta}) = \log p(\mathcal{D}_n \mid b_i, \boldsymbol{\theta})$, with maximum likelihood estimates $\widehat{\boldsymbol{\theta}}_r$. In a general likelihood setting, the selection of r (under H_A) can be based on a suitable information criterion. Among the different options, the AIC function emerges as the one that is probably most usually employed; this relies on minimizing the Kullback-Leibler (KL) divergence[22] between the correct and the fitted function. Specifically, for each $r \in \{0, 1, \ldots\}$, the AIC function is written as $\text{AIC}(r) = -2\mathcal{L}_r + 2(n_0 + r)$, where $\mathcal{L}_r$ is equal to the supremum of the log-likelihood taken over all the possible values for $\boldsymbol{\theta}$, $\mathcal{L}_r = \sup_{\boldsymbol{\theta}} L_r(\boldsymbol{\theta})$, while the second term prevents overfitting by penalizing the complexity of the model; particularly, n_0 stands for the total number of parameters in the SP joint model under H_0, and r denotes the number of additional basis terms adopted in H_A. To check H_0 against H_A, we compute $\Delta\text{AIC}(r) = \text{AIC}(r) - \text{AIC}(0) = 2(\mathcal{L}_r - \mathcal{L}_0) - 2r$, then rejecting H_0 whenever there exists an integer $r^* \in \{1, \ldots, R_n\}$ that maximizes $\Delta\text{AIC}(r^*) > 0$, where the bound R_n is theoretically expected to increase with n. Despite the widespread use of the AIC function as a model selection criterion, it is known that $\Pr(r^* = 0 \mid H_0) \longrightarrow 0.712$ as $n \longrightarrow \infty$,[23] yielding a test with asymptotic level $\alpha = 0.288$. Hence, one may conclude that a test based on the AIC overestimates the probability of selecting H_0 according to the levels most used in practice ($\alpha \leq 0.10$), while the test has little power of detecting nonlinear shapes truly accommodating $\widetilde{w}_i$. Following earlier work by Eubank and Hart[14] to circumvent this drawback, a modified AIC function is proposed in Aerts et al[16] by means of

$$\text{MAIC}(r, C_{n,\alpha}) = -2\mathcal{L}_r + C_{n,\alpha}(n_0 + r),\ r \geq 0, \tag{5}$$

where the constant $C_{n,\alpha} > 1$, which is related to both the sample size and the desired α-level of the test, penalizes higher significance levels more severely. Hence, unlike the traditional AIC, for which the penalty is fixed at 2 using theoretic-information considerations, the MAIC-based test permits a wide range of nominal levels ($\alpha \leq 0.288$) under H_0 once $C_{n,\alpha}$ is adequately selected. In practice, H_0 can be rejected whenever the order for which $\text{MAIC}(r, C_{n,\alpha})$ is maximum, say $r_C^* = \text{argmax}_{\,0 \leq r \leq R_n} \text{MAIC}(r, C_{n,\alpha})$, is larger than zero:

$$\text{reject } H_0 \text{ if } \Delta\text{MAIC}(r_C^*, C_{n,\alpha}) = 2(\mathcal{L}_r - \mathcal{L}_0) - C_{n,\alpha}\, r_C^* \geq 0 \text{ , for some } r_C^* \in \{1, \ldots, R_n\}. \tag{6}$$

This test can be reformulated by means of the test statistic $T_n = \max_{1 \leq r \leq R_n} 2(\mathcal{L}_r - \mathcal{L}_0)/r$, in terms of which the following decision rule for rejecting H_0 holds:

$$\text{reject } H_0 \text{ whenever } T_n \geq C_{n,\alpha}. \tag{7}$$

Using combinatorial results of Spitzer[25], a numerical solution for $C_{n,\alpha}$ can be uniquely computed at an asymptotic α-level from numerically solving

$$\exp\left\{ -\sum_{r=1}^{\infty} \frac{\Pr\left(\chi_r^2 > r\, C_{n,\alpha}\right)}{r} \right\} = 1 - \alpha, \tag{8}$$

where χ^2_r is a central chi-squared random variable with $r \geq 1$ degrees of freedom. Since $\Pr(T_n \leq x \mid \mathrm{H}_0) \longrightarrow 1 - \alpha$ as R_n and n tend to ∞, it is clear that the alternative version of the test yields an estimate for the true p-value can be derived for a given T_n.

For practical purposes, however, the requirement $R_n \longrightarrow \infty$ is not necessary to construct a valid testing procedure in terms of significance level and power. In fact, all that is needed is to choose a sufficiently large integer R_n. By proceeding this way, the test allows for using completely parametric estimators to properly approximate the infinite dimensional truth, ultimately checking the null function with nonparametric regression techniques. Different values have been proposed for R_n in the literature. Aerts et al[16] state that fixing $R_n = 10$ ensures a consistent test against all nonparametric alternatives. Zhang and Davidian[26] suggest using $R_n = 4$ in the context of semi-nonparametric linear mixed models. More recently, Nysen et al[17] extended their parametric base function by including, at most, $R_n = 7$ terms in the presence of censored data. Hereafter, this last suggestion is followed.

3 | MAXIMUM LIKELIHOOD ESTIMATION

The unknown parameters of the joint model are estimated through a likelihood-based approach, wherein the complete log-likelihood function of both the longitudinal and survival outcomes need to be explicit. Recovering the notation introduced in Section 3, let $\mathcal{D}_n$ and $\boldsymbol{\theta}$ denote the observed data and the full vector of estimable parameters, respectively. We assume that, given the time-independent random effects $\mathbf{b}_i$ for the ith subject, the longitudinal and time-to-event outcomes are independent, as are the n_i discrete set of longitudinal responses observed for each subject in the study. These conditional independence assumptions are formally postulated by

$$\begin{aligned} p\{\mathbf{y}_i, T_i, \delta_i \mid \mathbf{b}_i, \boldsymbol{\theta}, f(\cdot)\} &= \int_{-\infty}^{\infty} p_y(\mathbf{y}_i \mid \mathbf{b}_i, \boldsymbol{\theta}_y)\, p_t\{T_i, \delta_i \mid \mathcal{M}_i(t), \boldsymbol{\theta}_t\}\, p_b(\mathbf{b}_i \mid \boldsymbol{\theta}_b)\, \mathrm{d}\mathbf{b}_i, \\ p_y(\mathbf{y}_i \mid \mathbf{b}_i, \boldsymbol{\theta}_y) &= \prod_{j=1}^{n_i} p_y\{y_i(t_{ij}) \mid \mathbf{b}_i, \boldsymbol{\theta}_y\}, \end{aligned} \tag{9}$$

where $p_y(\cdot)$ is the conditional density function for the longitudinal measurements, $p_t(\cdot)$ is the conditional density for the event times, and $p_b(\cdot)$ is the conditional density for the normally-distributed random effects. The reader is referred to Rizopoulos[10] for the specific form of the probability density functions $p_y(\cdot)$, $p_t(\cdot)$, and $p_b(\cdot)$, as well as for some computational drawbacks that may arise in a maximum likelihood setting. By (9), the complete conditional joint log-likelihood function for $\boldsymbol{\theta}$ is satisfactorily expressed by a factorization akin to a shared random-effect parameterization:

$$\log p\{\mathcal{D}_n \mid \mathbf{b}_i, \boldsymbol{\theta}\} = \sum_{i=1}^{n} \log \int_{-\infty}^{\infty} \left[\prod_{j=1}^{n_i} p_y\{y_i(t_{ij}) \mid \mathbf{b}_i, \boldsymbol{\theta}_y\} \right] p_t\{T_i, \delta_i \mid \mathcal{M}_i(t), \boldsymbol{\theta}_t\}\, p_b(\mathbf{b}_i \mid \boldsymbol{\theta}_b)\, \mathrm{d}\mathbf{b}_i. \tag{10}$$

The maximum log-likelihood estimates for the parameters of the proposed SP joint model are obtained by maximizing (10) with respect to $\boldsymbol{\theta}$, resulting in the solution $\widehat{\boldsymbol{\theta}}$. Nevertheless, such a function does not have a computationally simple form, and its maximization implies the implementation of both numerical integration techniques and optimization algorithms. More specifically, the random terms can be assumed to be missing data, requiring an expectation-maximization (EM) algorithm.

4 | PREDICTIVE PERFORMANCE

Once a joint model has been fitted, the next step is to evaluate its accuracy in predicting time-to-event outcomes. To introduce this concept, consider a generic individual $i \in \{1, \ldots n\}$ from our target population, providing a history of CD4 responses up to time t, $\mathcal{Y}_i(t) = \{y_i(t_{ij}), 0 \leq t_{ij} \leq t\}$. Also, consider a set of fixed covariates $\mathbf{w}_i$ that are linearly related to the log-hazard, along with a single covariate $\widetilde{w}_i$ entering the model under an unknown shape. From the standard Bayesian theory and using (9), the probability of not reaching immunity over a time horizon $u = t + \Delta t$, given that immunity has not been reached at t, is:

$$\begin{aligned} \pi_i(u \mid t) &= \Pr(T_i^* \geq u \mid T_i^* > t, \mathcal{Y}_i(t), \mathbf{w}_i, \widetilde{w}_i, \mathcal{D}_n) \\ &= \int_{\boldsymbol{\theta}} \Pr(T_i^* \geq u \mid T_i^* > t, \mathcal{Y}_i(t), \mathbf{w}_i, \widetilde{w}_i, \mathcal{D}_n)\, p(\boldsymbol{\theta} \mid \mathcal{D}_n)\, \mathrm{d}\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\theta}} \int_{\mathbf{b}_i} \frac{\mathcal{S}_i\{u \mid \mathcal{M}(u, \mathbf{b}_i), \boldsymbol{\theta})\}}{\mathcal{S}_i\{t \mid \mathcal{M}(u, \mathbf{b}_i), \boldsymbol{\theta})\}}\, p(\mathbf{b}_i \mid T_i^* > t, \mathcal{Y}_i(t), \mathbf{w}_i, \widetilde{w}_i, \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \mathcal{D}_n)\, \mathrm{d}\mathbf{b}_i \mathrm{d}\boldsymbol{\theta}. \end{aligned} \tag{11}$$

Admitting a sufficiently large sample size, then $\boldsymbol{\theta} \sim \mathrm{N}\{\widehat{\boldsymbol{\theta}}, \widehat{\mathrm{var}}(\widehat{\boldsymbol{\theta}})\}$ with $\widehat{\mathrm{var}}(\widehat{\boldsymbol{\theta}}) = \{\mathcal{I}(\widehat{\boldsymbol{\theta}})\}^{-1}$. A Monte Carlo simulation scheme is then applied to ultimately obtain an estimate $\widehat{\pi}_i(u \mid t)$; further technical details can be found in Rizopoulos.[27] Of note, as new information is collected at a further time $t' > t$, the prediction obtained in (11) is dynamically updated to $\pi_i(u \mid t')$.

The predictive performance of the joint model is quantified by employing discrimination measures; that is, we focus on quantifying the ability of the fitted model to rank individuals from low to high risk. We follow Andrinopoulou[28] to explore the model's ability to correctly distinguish between subjects who are about to achieve immunity within a clinically relevant upcoming period $(t, u]$ after their last measurement and those who are not. More precisely, we assume that any generic individual i achieves immunity if $\pi_i(u \mid t) \leq c_D$, $c_D \in [0, 1]$, and the sensitivity (true positive rate) and specificity (true negative rate) associated are defined by $\mathrm{SN}(t, u) = \Pr\{\pi_i(u \mid t) \leq c_D \mid T^* \in (t, u]\}$ and $\mathrm{SP}(t, u) = \Pr\{\pi_i(u \mid t) > c_D \mid T^* > u\}$, respectively. The plot of $\mathrm{SN}(t, u)$ against $1 - \mathrm{SP}(t, u)$ for any cut-off value c_D yields the corresponding receiver operating characteristic curve, $\mathrm{ROC}(t, u)$, which provides graphical assessment of the overall discrimination capability. That is, for any randomly-chosen pair of subjects i_1 and i_2 such that subject i_1 experiences the event within $(t, u]$ whereas subject i_2 does not, the curve assess on whether subject i_2 effectively has a higher survival probability. This concordance probability can be summarized in a time-dynamic manner by the area under the ROC curve,

$$\mathrm{AUC}(t, u) = \Pr\left[\pi_{i_1}(u \mid t) < \pi_{i_2}(u \mid t) \mid \{T^*_{i_1} \in (t, u]\} \cap \{T^*_{i_2} > u\}\right], \tag{12}$$

where the higher the AUC is above 0.5, the better our SP joint model performs in predicting time to immune reconstitution. Because the longitudinal history is affected by censoring, the expression above can be estimated through its decomposition into four terms as

$$\widehat{\mathrm{AUC}}(t, u) = \sum_{d=1}^{4} \left[\frac{\sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} I\{\widehat{\pi}_{i_1}(u \mid t) < \widehat{\pi}_{i_2}(u \mid t)\} \times I\{\Omega_d(t, u)\} \times \widehat{W}_d}{\sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} I\{\Omega_d(t, u)\} \times \widehat{W}_d} \right],$$

where each contributing term, $d \in \{1, 2, 3, 4\}$, considers a set of subjects $\Omega_d(t)$ whose observed times $T_{i_1} = \min\{T^*_{i_1}, C_{i_1}\}$ and $T_{i_2} = \min\{T^*_{i_2}, C_{i_2}\}$ either can be ordered, as in the set $\Omega_1(t, u) = \{T^*_{i_1} \in (t, u]\} \cap \{T_{i_2} > u\}$, or cannot be ordered due to censoring, as in the remaining three sets: $\Omega_2(t, u) = \{C_{i_1} \in (t, u]\} \cap \{T_{i_2} > u\}$, $\Omega_3(t, u) = \{T^*_{i_1} \in (t, u]\} \cap \{T_i < T_{i_1} \leq u\}$, and $\Omega_4(t, u) = \{C_{i_1} \in (t, u]\} \cap \{T_{i_1} < C_{i_2} \leq u\}$. All these sets, in turn, are assigned a estimated censoring weight $\widehat{W}_d$ which accounts for the probability that times T_{i_1} and T_{i_2} are comparable; the first set leads to the trivial weight $\widehat{W}_1 = 1$, while the remaining sets match with $\widehat{W}_2 = 1 - \widehat{\pi}_{i_1}(u \mid T_{i_1})$, $\widehat{W}_3 = \widehat{\pi}_{i_2}(u \mid T_{i_2})$, and $\widehat{W}_4 = \{1 - \widehat{\pi}_{i_1}(u \mid T_{i_1})\}\widehat{\pi}_{i_2}(u \mid T_{i_2})$.

5 | SIMULATION STUDY

5.1 | General design

An extensive simulation study is designed to assess the performance of the MAIC-based test[16] in the context of SP joint models. Particularly, the interest lies in testing the underlying shape of a misspecified covariate. For this purpose, the ith subject's log-hazard for an event is postulated to depend on: a baseline log-hazard function, an intermittently collected time-dependent biomarker, a binary covariate w_i, and a nonlinear covariate $\widetilde{w}_i \sim U(0, 1)$. Five sample sizes are considered to examine the sample performance, namely $n \in \{250, 500, 750, 1000, 1500\}$ subjects monitored over a 15-year period, with each individual observed at up to $n_i = 5$ sequential time points which always include $t = 0$. Two different SP joint models are concretely specified as

$$\mathrm{JM_L}: \log h_i\{t \mid \mathcal{M}_i(t), w_i, \widetilde{w}_i\} = \log h_0(t) + \lambda\, \mu_i(t) + \gamma_w w_i + f(\widetilde{w}_i), \quad t > 0,$$

$$\mathrm{JM_S}: \log h_i\{t \mid \mathcal{M}_i(t), w_i, \widetilde{w}_i\} = \log h_0(t) + \lambda_1\, \mu_i(t) + \lambda_2 \frac{\mathrm{d}\mu_i(t)}{\mathrm{d}t} + \gamma_w w_i + f(\widetilde{w}_i), \quad t > 0,$$

where both models explicitly assume a biparametric Weibull law for the baseline hazard function. For $\mathrm{JM_L}$, longitudinal trajectories are modeled using correlated random-intercept and random-slope terms, $\mu_i(t) = \beta_0 + b_{i0} + (\beta_1 + b_{i1})\, t$. The latent structure is specified as $\mathbf{b}_i = (b_{i0}, b_{i1})^\top \sim \mathrm{N}_2(\mathbf{0}, \mathbf{D})$, wherein the 2×2 covariance matrix is $\mathbf{D} = \left[d_{11} = \sigma_0^2,\ d_{12} = \mathrm{cov}(b_0, b_1),\ d_{21} = \mathrm{cov}(b_1, b_0),\ d_{22} = \sigma_1^2\right]$, with $\mathrm{cov}(b_0, b_1) = \rho_{01}\sigma_0\sigma_1$ including the correlation term ρ_{01}. Furthermore, $\mathrm{JM_L}$ postulates that $\mathbf{g} \equiv g(\cdot) = \mathrm{Id}(\cdot)$, so $\mu_i(t)$ is additively related to the log-hazard for an event. On the other hand, when defining $\mathrm{JM_S}$, longitudinal profiles are allowed to be flexibly accounted for via natural cubic B-splines of time, $\mu_i(t) = \beta_0 + b_{i0} + (\beta_{N_1} + b_{iN_1}) N_{3,1}(t, \xi_{\mathrm{med}}) +$

$(\beta_{N_2} + b_{iN_2})N_{3,2}(t, \xi_{\text{med}}) + \beta_w w_i$, with a single internal knot ξ_{med} placed at the median of time measurements. The random effects are assumed to be independent, being encompassed by the 3×3 covariance matrix $\mathbf{D} = \text{diag}\left[d_{11} = \sigma_0^2, d_{22} = \sigma_{N_1}^2, d_{33} = \sigma_{N_2}^2\right]$, with $b_{i0} \sim \text{N}(0, \sigma_0^2)$, $b_{iN_1} \sim \text{N}(0, \sigma_{N_1}^2)$ and $b_{iN_2} \sim \text{N}(0, \sigma_{N_2}^2)$. Also, as an alternative g-parameterization, JM_S is defined through $\mathbf{g} = \{g_1(\cdot), g_2(\cdot)\}^\top = \{\text{Id}(\cdot), \text{d}(\cdot)/\text{d}t\}^\top$, assuming that the conditional log-hazard is additively related to both $\mu_i(t)$ and its rate of change; this defines the association vector $\lambda = (\lambda_1, \lambda_2)^\top$. Both JM_L and JM_S are combined with two levels of random right censoring: 25% (moderate) and 45% (strong). These levels are imposed by the generation of independent, uniform random variables on the $[0, k_{25}(n)]$ and $[0, k_{45}(n)]$ intervals, respectively, with the n-dependent upper bounds being suitably chosen.

Depending on the test feature to be examined, the true function $f(\cdot)$ embracing $\widetilde{\omega}_i$ is chosen as follows. To evaluate the robustness feature, $f(\cdot)$ is simulated under $\mathcal{F}_0$, i.e., $f(\widetilde{\omega}_i) = c_1\widetilde{\omega}_i$ with $c_1 \neq 0$. On the other hand, when providing insight into the power of the test to recognize false null hypotheses, $f(\cdot)$ is truly modeled under H_A. More specifically, the functional shapes adopted are $f_{\text{POL}}(\widetilde{\omega}_i) = (c_1 - c_2)\widetilde{\omega}_i + c_2\widetilde{\omega}_i^2$ and $f_{\text{SIN}}(\widetilde{\omega}_i) = c_1\widetilde{\omega}_i + c_2\sin(\pi\widetilde{\omega}_i)$, wherein larger absolute values for $c_2 \neq 0$ are expected to indicate higher departures from H_0. Next, to examine the performance of the two aforementioned test features, the linearity between the conditional log-hazard function and $\widetilde{\omega}_i$ needs to be checked. This is carried out by estimating $f(\cdot)$ through a nonparametric estimator $\varphi(\cdot)$, easily constructed from two possible sets of basis functions that expand the linear model assumed in H_0: orthogonal polynomials, defined by $\{u_r(\widetilde{\omega}_i) = \widetilde{\omega}_i^{1+r}\}$, and a set of sine functions of the form $\{u_r(\widetilde{\omega}_i) = \sin(\pi r\widetilde{\omega}_i),\ A > 0\}$. Thus, $f(\cdot)$ is approximated by either $\varphi_{\text{POL}}(\widetilde{\omega}_i \mid \mathbf{c}) = c_1\widetilde{\omega}_i + \sum_{k=1}^{r} c_{1+k}\widetilde{\omega}_i^{1+k}$ or $\varphi_{\text{SIN}}(\widetilde{\omega}_i \mid \mathbf{c}) = c_1\widetilde{\omega}_i + \sum_{k=1}^{r} c_{1+k}\sin(\pi k\widetilde{\omega}_i)$, respectively. In both series expansions, the order is determined by the maximum r over $\{0, 1, \ldots, 7\}$ for which H_0 is rejected.

Table 1 summarizes all possible simulation scenarios considered to examine the MAIC-based test performance in a SP joint modeling context. When assessing the robustness of the test, JM_L and JM_S are combined with the 25% and 45% censoring levels, together with the nonparametric estimators $\varphi_{\text{POL}}(\cdot)$ and $\varphi_{\text{SIN}}(\cdot)$, adopted to expand H_0. This results in a total of $2 \times 2^2 = 8$ scenarios. Likewise, when examining the power of the test, the two SP joint models are reconsidered at the two aforementioned censoring levels, but now the true underlying shape of $\widetilde{\omega}_i$ is either $f_{\text{POL}}(\cdot)$ or $f_{\text{SIN}}(\cdot)$. These true shapes are again approximated by either $\varphi_{\text{POL}}(\cdot)$ or $\varphi_{\text{SIN}}(\cdot)$, yielding $4 \times 2^3 = 32$ scenarios. Within a given scenario (either to assess the robustness or power of the test) and sample size, a number $M = 500$ independent datasets is simulated. Then, for a given dataset, the linearity of $\widetilde{\omega}_i$ (order $r = 0$) in a particular SP joint model is tested by considering the set $r \in \{1, \ldots, 7\}$ of alternative orders provided by $\varphi(\cdot)$. As a consequence, assessing the robustness of the test comprises the fitting of $8 \times 5 \times 8 \times 500 = 160,000$ SP joint models, while evaluating the power of the test entails $32 \times 5 \times 8 \times 500 = 640,000$ SP joint models.

TABLE 1 Simulation scenarios to be accounted for when assessing the MAIC-based test performance.

Assessment	Model	Cens.	Truth	c_2	Estim.	Scenarios
Robustness	JM_L	{25%, 45%}	$f \in \mathcal{F}_0$	–	$\{\varphi_{\text{POL}}, \varphi_{\text{SIN}}\}$	2^2
	JM_S	{25%, 45%}	$f \in \mathcal{F}_0$	–	$\{\varphi_{\text{POL}}, \varphi_{\text{SIN}}\}$	2^2
Power	JM_L	{25%, 45%}	f_{POL}	{1.50, 2.00}	$\{\varphi_{\text{POL}}, \varphi_{\text{SIN}}\}$	2^3
			f_{SIN}	{0.50, 0.75}	$\{\varphi_{\text{POL}}, \varphi_{\text{SIN}}\}$	2^3
	JM_S	{25%, 45%}	f_{POL}	{2.00, 2.50}	$\{\varphi_{\text{POL}}, \varphi_{\text{SIN}}\}$	2^3
			f_{SIN}	{0.50, 0.75}	$\{\varphi_{\text{POL}}, \varphi_{\text{SIN}}\}$	2^3

For the longitudinal process, the true parameter values in JM_L are $(\beta_0 = 7.20, \beta_1 = -0.38)^\top$, with $d_{11} = 3.50$, $d_{12} = -0.10$, and $d_{22} = 0.10$. For JM_S, the true values are $(\beta_0 = 7.10, \beta_{N_1} = -3.80, \beta_{N_2} = -3.25)^\top$, with $d_{11} = 4.50$, $d_{22} = 3.50$, and $d_{33} = 2.75$. In each of these two possible specifications, $\sigma_\varepsilon^2 = 1.50^2$. For the survival process, datasets simulated from JM_L have $\gamma_w = 0.18$, $\lambda = -0.20$, $c_1 = 0.10$, while the values for parameter c_2 are either $\{1.50, 2.00\}$ when $f_{\text{POL}}(\cdot)$ is the true function for $\widetilde{\omega}_i$, or $\{0.50, 0.75\}$ when $f_{\text{SIN}}(\cdot)$ truly holds. Furthermore, the true parameter values for $h_0(\cdot)$ are either $\nu = 0.075$ and $\sigma_t = 1.385$, for 25% censoring, or $\nu = 0.022$ and $\sigma_t = 1.590$, for 45% censoring. The true parameter setting for JM_S is established to be $\gamma_w = 0.150$, $(\lambda_1 = -0.250, \lambda_2 = 1.450)^\top$, and $c_1 = 3.500$, while $h_0(\cdot)$ is simulated with either $\nu = 0.054$ and $\sigma_t = 1.110$ or $\nu = 0.007$ and $\sigma_t = 1.560$, for 25% or 45% censoring, respectively. Appendix B reports the results from the simulations designed to show how our algorithm performs for the maximum likelihood estimates (Tables S3 to S6). These results, respectively corresponding to each possible combination of JM_L or JM_S at a certain censoring level, indicate that the estimates $\widehat{\boldsymbol{\theta}}$ approach the true values $\boldsymbol{\theta}$ as n becomes larger over the sequence $\{250, 500, 750, 1000, 1500\}$. In parallel, the different error measurements decrease in the large-sample limit.

5.2 | Information computed and simulation results

Robustness of the test

The asymptotic α-level of the MAIC-based test is carefully assessed by estimating the derived $C_{n,\alpha}$ when H_0 is likely to be true. Within a particular scenario, size n, and the mth replicated dataset, $m = 1, \ldots, M$, the interest consists of testing whether an order $r \in \{1, \ldots, 7\}$ of $\varphi(\cdot)$ outperforms the linear accommodation of $\widetilde{\omega}_i$ when a given SP joint model is fitted to the dataset. Over the bounded support for r, the mth observed test statistic $T_n^{(m)} = \max_{1 \leq r \leq 7} \left\{ 2(\mathcal{L}_r^{(m)} - \mathcal{L}_0^{(m)})/r \right\}$ is computed, and its value is retained whenever $r_c^* > 1$ leads to $T_n^{(m)} > 0$. Subsequently, from (7), the α-level for H_0 can be indirectly evaluated through the agreement between the $(1-\alpha)$ empirical quantile over the set $\left\{ T_n^{(1)}, \ldots, T_n^{(M)} \right\}$, say $q_{n,\alpha} \equiv \widehat{C}_{n,\alpha}$, and the theoretical penalty $C_{n,\alpha}$. Occasionally, the mth maximum likelihood estimation for a given parameter setting may be affected by numerical instabilities, yielding either extremely large or negative values for $T_n^{(m)}$. To avoid biased estimates, this small fraction of anomalous responses is omitted when computing $\widehat{C}_{n,\alpha}$. The assessment of the test significance level is undertaken for nominal values $\alpha = 0.05$ and $\alpha = 0.10$, through their corresponding theoretical penalties $C_{n,0.05} = 4.179$ and $C_{n,0.10} = 3.221$, resulting from (8). This testing procedure has the added value of providing guidance as regards the relationship between increasing sample sizes and the bound R_n needed to obtain r_c^*, ultimately informing whether the range of order values considered gives enough evidence to suggest that no significant improvement is to be had from trying higher orders. The simulation results in Table 2 provide, at each α-level and scenario, the estimated penalties $\widehat{C}_{n,\alpha}$ obtained across the five different sample sizes considered. Of interest, these empirical penalties converge rapidly to the theoretical ones with an increasing n, even for the smallest size $n = 250$. That is, $\Pr(T_n > \widehat{C}_{n,\alpha} \mid H_0) \longrightarrow \alpha$ as $n \longrightarrow \infty$. This indicates that, under the joint modeling framework, the level-robustness of the test behaves as described in theory. Further, the Type I error probability does not seem to be substantially connected either to the choice of the censoring level or with the estimator by which the null function is expanded. This conclusion is also graphically assessed in Section C1 from Appendix C (Figures S2 to S9). Therein, the empirically derived density functions of T_n under H_0 are depicted for each scenario and n, while in each case plotting the theoretical penalty terms $C_{n,\alpha}$ at the prescribed levels of significance, $\alpha \in \{0.05, 0.10\}$.

TABLE 2 Empirical penalties of the MAIC-based test from the simulation study.

				$\widehat{C}_{n,0.05}$ for a given n					$\widehat{C}_{n,0.10}$ for a given n				
Model	**Cens.**	**Truth**	**Estim.**	**250**	**500**	**750**	**1000**	**1500**	**250**	**500**	**750**	**1000**	**1500**
JM_L	25%	$f \in \mathcal{F}_0$	φ_{POL}	4.382	4.407	4.176	4.022	4.108	3.302	3.157	3.208	3.175	3.246
			φ_{SIN}	4.334	4.360	4.100	4.045	4.077	3.351	3.199	3.059	3.057	3.229
	45%	$f \in \mathcal{F}_0$	φ_{POL}	4.420	4.536	4.048	4.056	4.204	3.236	3.643	3.083	3.287	3.248
			φ_{SIN}	4.392	4.516	4.003	4.208	4.147	3.410	3.465	2.932	3.191	3.145
JM_S	25%	$f \in \mathcal{F}_0$	φ_{POL}	4.241	4.028	4.164	4.088	4.229	3.362	3.277	3.154	3.064	3.293
			φ_{SIN}	4.136	4.173	3.844	4.089	4.091	3.265	3.120	3.023	3.074	3.320
	45%	$f \in \mathcal{F}_0$	φ_{POL}	4.532	3.900	3.998	4.216	4.061	3.492	3.157	3.110	3.423	3.129
			φ_{SIN}	4.411	4.258	4.107	4.204	3.961	3.405	3.344	3.117	3.320	3.282

In addition, it is worthwhile to seek some further guidance on the traditionally assumed increasing dependence between the maximizer $r_c^* > 1$ of the MAIC-based test and n, at a prespecified α-level. In general, the lower the order to consistently estimate $C_{n,\alpha}$ within each scenario, the lower the value needed for the bound R_n, yielding important computational savings. As noted above, we select $R_n = 7$ as a default upper order to be tested in all scenarios. Departing from this initial reference, we can explore whether it is possible to effectively reduce the set of order candidates to be considered. Section C2 of Appendix C provides, for a particular alternative and n, the $r_c^* \in \{1, \ldots, 7\}$ for which the maximum value of $C_{n,\alpha}$ is attained at a given α (Tables S7 to S14). Examination of these tables reveals that for $r \geq 3$ the error in approaching $C_{n,\alpha}$ is relatively negligible in most cases. Thus, above $r = 3$, the theoretical significance level is no longer being approximated, so estimating further coefficients in $\mathbf{c}$ does not yield any substantial improvement in the practical performance of $\varphi(\cdot)$. This also aligns with the *weighted* nature of the log-likelihood ratio statistic T_n, providing the largest statistic among a set in which larger weights tend to be assigned at lower order alternatives; see, for example, Hart.[1]

Power of the test

Along with the assessment of the asymptotic level, we also provide insight into the power of the test to recognize false null hypotheses. Working from the SP joint models given by JM_L and JM_S, it is now supposed that the $\mathcal{F}_0$ class no longer holds to correctly include $\widetilde{\omega}_i$. Rather, this covariate is truly modeled under H_A, wherein $\widetilde{\omega}_i$ actually has one of these two possible true shapes: $f_{POL}(\cdot)$ or $f_{SIN}(\cdot)$. A stream of $m = 1, \ldots, M$ replicated MAIC-based tests are conducted under H_A for each of the experimental scenarios and sample sizes, while taking $\alpha = 0.05$ or $\alpha = 0.10$. For a specific scenario, the mth test statistic $T_n^{(m)}$ is once again that which maximizes the divergence from the null case over $r \in \{0, 1, \ldots, 7\}$. The empirical test power is computed utilizing the full proportion of times in which H_0 is correctly rejected for a given α and n, i.e., $\sum_{m=1}^{M} I(T_n^{(m)} \geq C_{n,\alpha})/M$. The results are presented in Table 3 and indicate that the power tends to 1 as n increases for a particular setting, reflecting the consistency of the procedure. Naturally, for a given $\alpha \in (0, 1)$, the larger c_2 is, the better the test detects that $\widetilde{\omega}_i$ is not correctly entered into the SP joint model under H_0, so convergence to 1 is faster.

TABLE 3 Empirical powers of the MAIC-based test from the simulation study.

					Power for $\alpha = 0.05$ and a given n					Power for $\alpha = 0.10$ and a given n				
Model	**Cens.**	**Truth**	c_2	**Estim.**	**250**	**500**	**750**	**1000**	**1500**	**250**	**500**	**750**	**1000**	**1500**
JM_L	25%	f_{POL}	1.50	φ_{POL}	0.256	0.500	0.636	0.774	0.940	0.382	0.604	0.736	0.860	0.960
				φ_{SIN}	0.260	0.500	0.638	0.776	0.942	0.378	0.600	0.734	0.852	0.960
			2.00	φ_{POL}	0.426	0.720	0.858	0.966	0.992	0.552	0.796	0.918	0.978	0.998
				φ_{SIN}	0.428	0.720	0.854	0.964	0.992	0.554	0.792	0.918	0.976	0.998
		f_{SIN}	0.50	φ_{POL}	0.510	0.800	0.936	0.992	0.998	0.620	0.850	0.964	0.994	0.998
				φ_{SIN}	0.516	0.806	0.932	0.986	0.998	0.628	0.848	0.962	0.994	0.998
			0.75	φ_{POL}	0.856	0.988	1.000	1.000	1.000	0.926	0.992	0.996	1.000	1.000
				φ_{SIN}	0.864	0.990	1.000	1.000	1.000	0.930	0.994	1.000	1.000	1.000
	45%	f_{POL}	1.50	φ_{POL}	0.184	0.380	0.486	0.616	0.792	0.260	0.484	0.598	0.720	0.880
				φ_{SIN}	0.186	0.378	0.484	0.620	0.798	0.264	0.488	0.600	0.728	0.876
			2.00	φ_{POL}	0.312	0.580	0.728	0.844	0.960	0.410	0.674	0.818	0.892	0.986
				φ_{SIN}	0.312	0.580	0.728	0.840	0.960	0.412	0.680	0.824	0.892	0.982
		f_{SIN}	0.50	φ_{POL}	0.426	0.668	0.862	0.948	0.992	0.544	0.746	0.922	0.974	0.996
				φ_{SIN}	0.428	0.662	0.868	0.952	0.994	0.548	0.752	0.918	0.972	0.996
			0.75	φ_{POL}	0.774	0.952	0.996	1.000	1.000	0.870	0.974	0.996	1.000	1.000
				φ_{SIN}	0.782	0.952	0.996	1.000	1.000	0.876	0.970	0.996	1.000	1.000
JM_S	25%	f_{POL}	2.00	φ_{POL}	0.463	0.710	0.865	0.925	0.995	0.565	0.805	0.905	0.950	1.000
				φ_{SIN}	0.458	0.699	0.861	0.929	0.998	0.563	0.802	0.914	0.960	0.997
			2.50	φ_{POL}	0.561	0.859	0.956	0.974	1.000	0.647	0.909	0.971	0.990	1.000
				φ_{SIN}	0.585	0.853	0.965	0.974	0.997	0.659	0.917	0.971	0.984	1.000
		f_{SIN}	0.50	φ_{POL}	0.325	0.550	0.712	0.835	0.949	0.439	0.641	0.815	0.888	0.969
				φ_{SIN}	0.326	0.554	0.702	0.830	0.950	0.437	0.639	0.816	0.897	0.969
			0.75	φ_{POL}	0.654	0.893	0.977	0.990	1.000	0.729	0.934	0.987	0.993	1.000
				φ_{SIN}	0.658	0.895	0.967	0.993	1.000	0.742	0.938	0.985	0.995	1.000
	45%	f_{POL}	2.00	φ_{POL}	0.327	0.548	0.708	0.803	0.938	0.443	0.638	0.793	0.881	0.966
				φ_{SIN}	0.312	0.542	0.709	0.812	0.950	0.432	0.645	0.789	0.876	0.964
			2.50	φ_{POL}	0.406	0.701	0.854	0.910	0.995	0.485	0.763	0.904	0.947	1.000
				φ_{SIN}	0.383	0.712	0.794	0.928	0.988	0.490	0.787	0.870	0.962	1.000
		f_{SIN}	0.50	φ_{POL}	0.245	0.339	0.505	0.582	0.786	0.347	0.436	0.611	0.691	0.861
				φ_{SIN}	0.221	0.367	0.461	0.619	0.752	0.295	0.471	0.561	0.714	0.808
			0.75	φ_{POL}	0.434	0.731	0.874	0.923	0.987	0.529	0.814	0.940	0.960	0.996
				φ_{SIN}	0.421	0.723	0.869	0.922	0.989	0.521	0.814	0.937	0.960	0.996

Discriminative performance

Another crucial issue involves quantifying, in terms of the predictive accuracy of the SP joint model, the implications of neglecting the nonlinear relationship between the log-hazard function and covariate $\widetilde{\omega}_i$. Taking advantage of part of the replications used to assess the power, we reconsider the $M = 500$ datasets simulated from JM_S with either 25% or 45% censoring and true function $f_{\text{POL}}(\widetilde{\omega}_i \mid \mathbf{c}) = (c_1 - c_2)\,\widetilde{\omega}_i + c_2\widetilde{\omega}_i^2$. Within each subscenario and $n \in \{750, 1500\}$, the term $\widehat{\text{AUC}}_r^{(m)}(t,u)$ is computed for $m = 1, \ldots, M$, while $f(\cdot)$ is approximated by $f_{\text{POL}}(\cdot)$ in these two possible situations: a linear estimator ($r = 1$) when assuming that H_0 holds, and a quadratic estimator ($r = 1$) under H_A. Then, the time-dependent variation of the discriminative performance for the mth dataset is $\Delta\widehat{\text{AUC}}^{(m)}(t,u) = \widehat{\text{AUC}}_1^{(m)}(t,u) - \widehat{\text{AUC}}_0^{(m)}(t,u)$, which yields an average improvement of $\Delta\widehat{\text{AUC}}_{\text{av}}^{(m)}(t,u) = \sum_{m=1}^{M} \widehat{\text{AUC}}^{(m)}(t,u)/M$.

Table 4 summarizes the estimated predictive measurements, where the follow-up periods $(t,u]$ are assumed to start at $t \in \{1.5, 2.5, 4.5\}$ and last one year, i.e. $u = t + 1$. Results report an improvement in the model's accuracy for each possible combination of censoring level, parameter c_2 and time t when sample size increases. Roughly, the increase in the overall discrimination capability ranges from 0.039 to 0.058 points across the different cases, entailing an improvement between 5.1% and 7.3% for the corresponding $\widehat{\text{AUC}}(t,u)$. A marginal improvement is confirmed as well when the quadratic effect $c_2 = 2$ is replaced by $c_2 = 2.5$, as expected. In parallel, as an additional intriguing result, the estimates are shown to behave robustly with different degrees of censoring in a certain subscenario.

TABLE 4 Simulation results for the discriminative performance when assuming JM_S with a quadratic polynomial function to truly accommodate $\widetilde{\omega}_i$. For both H_0 and H_A, the AUC is estimated at medically relevant intervals $[t, t+1]$, $t \in \{1.5, 2.5, 4.5\}$, using in each case 500 simulated datasets with $n \in \{750, 1500\}$ subjects and either 25% or 45% censoring. The variation in the AUC estimates between H_A ($r = 1$) and H_0 ($r = 0$) is computed, as well as the percentage improvement in the accuracy of the correct functional shape over the null model.

				$n = 750,\ u = t + 1$			$n = 1500,\ u = t + 1$		
Cens.	**Truth**	c_2	t	$\widehat{\text{AUC}}_0(t,u)$	$\widehat{\text{AUC}}_1(t,u)$	$\Delta\widehat{\text{AUC}}(t,u)$	$\widehat{\text{AUC}}_0(t,u)$	$\widehat{\text{AUC}}_1(t,u)$	$\Delta\widehat{\text{AUC}}(t,u)$
25%	f_{POL}	2.00	1.50	0.774	0.816	0.042 (5.5%)	0.780	0.824	0.044 (5.7%)
			2.50	0.778	0.831	0.053 (6.8%)	0.785	0.840	0.055 (7.0%)
			4.50	0.770	0.818	0.048 (6.2%)	0.778	0.827	0.049 (6.3%)
		2.50	1.50	0.782	0.829	0.047 (6.0%)	0.789	0.837	0.048 (6.1%)
			2.50	0.784	0.840	0.056 (7.2%)	0.793	0.851	0.058 (7.3%)
			4.50	0.775	0.826	0.051 (6.6%)	0.784	0.838	0.054 (6.9%)
45%	f_{POL}	2.00	1.50	0.756	0.797	0.041 (5.4%)	0.759	0.800	0.041 (5.4%)
			2.50	0.762	0.812	0.050 (6.5%)	0.764	0.816	0.052 (6.8%)
			2.50	0.758	0.803	0.045 (5.9%)	0.760	0.806	0.046 (6.0%)
		2.50	1.50	0.759	0.798	0.039 (5.1%)	0.762	0.802	0.040 (5.2%)
			2.50	0.764	0.816	0.052 (6.8%)	0.768	0.823	0.055 (7.2%)
			4.50	0.760	0.807	0.047 (6.1%)	0.761	0.809	0.048 (6.3%)

6 | ANALYSIS OF THE REAL DATA

The proposed methodology is illustrated by means of the previously introduced HIV data. Based on a given association structure, distinct SP joint models are employed to associate the subject-specific CD4 measurements and the time to immune recovery. Our ultimate goal is to ascertain whether there is sufficient evidence to reject a linear relationship between the log-hazard and the rescaled nadir CD4. In doing so, such a covariate is embedded in an rth order nonparametric estimator as in (4), this enabling us to test the linear assumption against a predefined alternative set of functional shapes. We select the SP joint model where the accommodation of nadir CD4 for a given $r \geq 1$, say r_c^*, provides stronger evidence for rejecting the linear shape ($r = 0$), and then compare the predictive abilities of the two hypotheses.

The monitoring period for the CD4 cell count has a median follow-up time of 2 years, with an average of 5.4 measurements per subject. To address the positively-skewed distribution of CD4 cells over time, our longitudinal approach accommodates an individual's measurements on the square root scale, $y_i(t) = \sqrt{\text{CD4}_i(t)},\ i = 1, \ldots, 628$. A flexible yet parametric specification of each subject trajectory is modeled via a mixed-effects model as follows. The fixed-effects part includes the intercept term along with the interaction between the sex indicator, equal to 1 for females, and the time effect, measured in years. The latter is included via natural cubic B-splines with a single internal knot ξ_{med}, allocated at the median of the follow-up times. The random-effects part includes a random term related to each of the B-spline basis functions. The syntax for the longitudinal process is

$$\begin{aligned} y_i(t) = \mu_i(t) + \varepsilon_i(t) = \beta_0 + \beta_{\text{s}\times N_1}\{\text{sex}_i \times N_{3,1}(t, \xi_{\text{med}})\} + \beta_{\text{s}\times N_2}\{\text{sex}_i \times N_{3,2}(t, \xi_{\text{med}})\} \\ + b_{i0} + b_{iN_1} N_{3,1}(t, \xi_{\text{med}}) + b_{iN_2} N_{3,2}(t, \xi_{\text{med}}) + \varepsilon_i(t), \end{aligned}$$

where the random terms are assumed to be independently distributed, $\mathbf{b}_i \sim \text{N}(0, \mathbf{D})$, with covariance matrix $\mathbf{D} = \text{diag}\left[\sigma_0^2, \sigma_{N_1}^2, \sigma_{N_2}^2\right]$. As a standard formulation for the survival approach, the log-hazard for immune recovery is directly related to the slope of the latent trajectory given by the square-root CD4 responses, while initially controlling for: the subject's gender $w_i = \text{sex}_i$ (0 = Male, 1 = Female), the nadir CD4 cell count rescaled to the unit interval $[0, 1]$, $\widetilde{w}_i = \text{nadCD4}_i$, and the interaction term between these two covariates.

For the survival process, previous HIV research has focused on the association, at a given t, between disease progression and the recorded value for the CD4 cell count; that is $\mathbf{g} \equiv g(\cdot) = \text{Id}(\cdot)$.[29,30] Seeking a more helpful portrait of reality, recent work advocates an alternative g-parameterization of the CD4 responses, based on the subject-specific rate of change in the biomarker pattern over time.[31] Here, the conditional log-hazard for immune recovery is related, at any time t, to the slope of the latent trajectory of square-root CD4 cells. With the aforesaid association structures in mind, three possible g-parameterizations of $\mu_i(t)$ are considered to be linked to the survival process: the current value, the slope, and both of these g-functions simultaneously. This leads to the corresponding three alternative SP joint models, wherein the possibly nonlinear effect of the nadir CD4 is allowed to be accommodated by an rth order nonparametric estimator $\varphi(\cdot)$. The latter, postulated as an alternative in order to expand the linear effect ($r = 0$) assumed in H_0 by $f(\cdot)$, leads us to examine a set of nested models under the MAIC-based testing procedure. When some degree of smoothness can be guaranteed for $f(\cdot)$, orthogonal polynomials of the type $\varphi_{\text{POL}}(\text{nadCD4}_i \mid \mathbf{c}) = c_1 \text{nadCD4}_i + \sum_{k=1}^{r} c_{1+k} \text{nadCD4}_i^{1+k},\ r \geq 1$, appear as a reasonably parsimonious option for expanding the linear function. As said above, the use of an orthogonal span allows us to compute uncorrelated regression coefficients, thus avoiding numerical instabilities in the estimation procedure.

The three SP joint models under consideration can be particularized as

$$\begin{aligned} \text{JM}_1 &: \log h_i\{t \mid \mathcal{M}_i(t), \text{sex}_i, \text{nadCD4}_i\} = \log h_0(t) + \lambda \mu_i(t) + \gamma_{\text{s}} \text{sex}_i \\ &\quad + \sum_{k=0}^{r} c_{1+k} \text{nadCD4}_i^{1+k} + \sum_{k=0}^{r} \gamma_{\text{int},1+k}\{\text{sex}_i \times \text{nadCD4}_i^{1+k}\}, \\ \text{JM}_2 &: \log h_i\{t \mid \mathcal{M}_i(t), \text{sex}_i, \text{nadCD4}_i\} = \log h_0(t) + \lambda \frac{\text{d}\mu_i(t)}{\text{d}t} + \gamma_{\text{s}} \text{sex}_i \\ &\quad + \sum_{k=0}^{r} c_{1+k} \text{nadCD4}_i^{1+k} + \sum_{k=0}^{r} \gamma_{\text{int},1+k}\{\text{sex}_i \times \text{nadCD4}_i^{1+k}\}, \\ \text{JM}_3 &: \log h_i\{t \mid \mathcal{M}_i(t), \text{sex}_i, \text{nadCD4}_i\} = \log h_0(t) + \lambda_1 \mu_i(t) + \lambda_2 \frac{\text{d}\mu_i(t)}{\text{d}t} + \gamma_{\text{s}} \text{sex}_i \\ &\quad + \sum_{k=0}^{r} c_{1+k} \text{nadCD4}_i^{1+k} + \sum_{k=0}^{r} \gamma_{\text{int},1+k}\{\text{sex}_i \times \text{nadCD4}_i^{1+k}\}, \end{aligned}$$

wherein the log-baseline hazard adopts a Weibull law. Within each alternative, the SP joint model accounting for the rth order estimator is compared to the corresponding null SP joint model by means of the smooth-based MAIC testing procedure.

Table 5 shows, for each of the three SP joint models, the corresponding set of log-likelihood ratio statistics computed over the sequence $1 \leq r \leq R_n = 7$. A specific value for test statistic T_n is then obtained for every model, with greater empirical evidence against the null hypothesis as T_n becomes larger. The maximum T_n is attained for JM_2 at $r_C^* = 1$, say $T_n = 3.934$, from which (7) and (8) allow us to estimate p-value $= 0.059$ when assuming a second-degree polynomial to incorporate the nadir count effect. Furthermore, the rescaled nadir CD4 is found to play a role in predicting immune recovery, $\widehat{c}_1 = 2.682$ (95% CI: 0.767, 4.597), while the slope of the square-root CD4 cell count trajectory is linked to the achievement of immune recovery, with $\widehat{\lambda} = 0.503$

(95% CI: 0.422, 0.584). Specifically, for subjects having the same sex and nadir CD4, the log-hazard ratio for a unit increase in the slope of the trajectory is exp (0.503) = 1.654 (95% CI: 1.525, 1.794).

TABLE 5 Results of applying the MAIC-based test to the HIV data for evaluating the linearity of $f(\cdot)$ in embracing the rescaled nadir CD4 within the SP joint modeling context. For each order $r \in \{1, \ldots, 7\}$, the corresponding log-likelihood ratio statistic $2(\mathcal{L}_r - \mathcal{L}_0)/r$ is computed, as well as the asymptotic α-level.

	JM_1 (value)		JM_2 (slope)		JM_3 (value + slope)	
r	$2(\mathcal{L}_r - \mathcal{L}_0)/r$	α	$2(\mathcal{L}_r - \mathcal{L}_0)/r$	α	$2(\mathcal{L}_r - \mathcal{L}_0)/r$	α
1	2.018	0.283	3.934	0.059	2.228	0.231
2	1.493	0.503	3.544	0.078	1.542	0.474
3	1.395	0.568	3.107	0.109	1.368	0.588
4	1.565	0.462	3.104	0.110	1.238	0.700
5	1.647	0.420	2.692	0.153	1.296	0.645
6	1.409	0.558	2.255	0.224	1.478	0.512
7	1.236	0.702	2.060	0.272	1.317	0.629

The predictive ability of the aforementioned model is evaluated by estimating the AUC(t, u) at follow-up times $t \in \{1.5, 3\}$ years, in conjunction with a 1.5-year medically relevant window, i.e., $u = t + 1.5$. To prevent overfitting in the AUC(t, u) estimates, we apply a 4-fold cross-validation procedure 50 times on our HIV data, so for the mth replication, $m = 1, \ldots, 50$, the 628 individuals are randomly divided into four distinct groups. The proposed joint models are then fitted to three out of the four groups, i.e. the training set, in each case assuming both a linear and a quadratic functional shape for the rescaled nadir CD4. The improvement between these two possible shapes is examined by computing $\Delta\widehat{\text{AUC}}^{(m)}(t, u) = \widehat{\text{AUC}}_1^{(m)}(t, u) - \widehat{\text{AUC}}_0^{(m)}(t, u)$ for the fourth group. Figure 2 displays these accuracy results, reporting a slightly better discriminative performance in all models when nadCD4_i is accommodated using a quadratic polynomial.

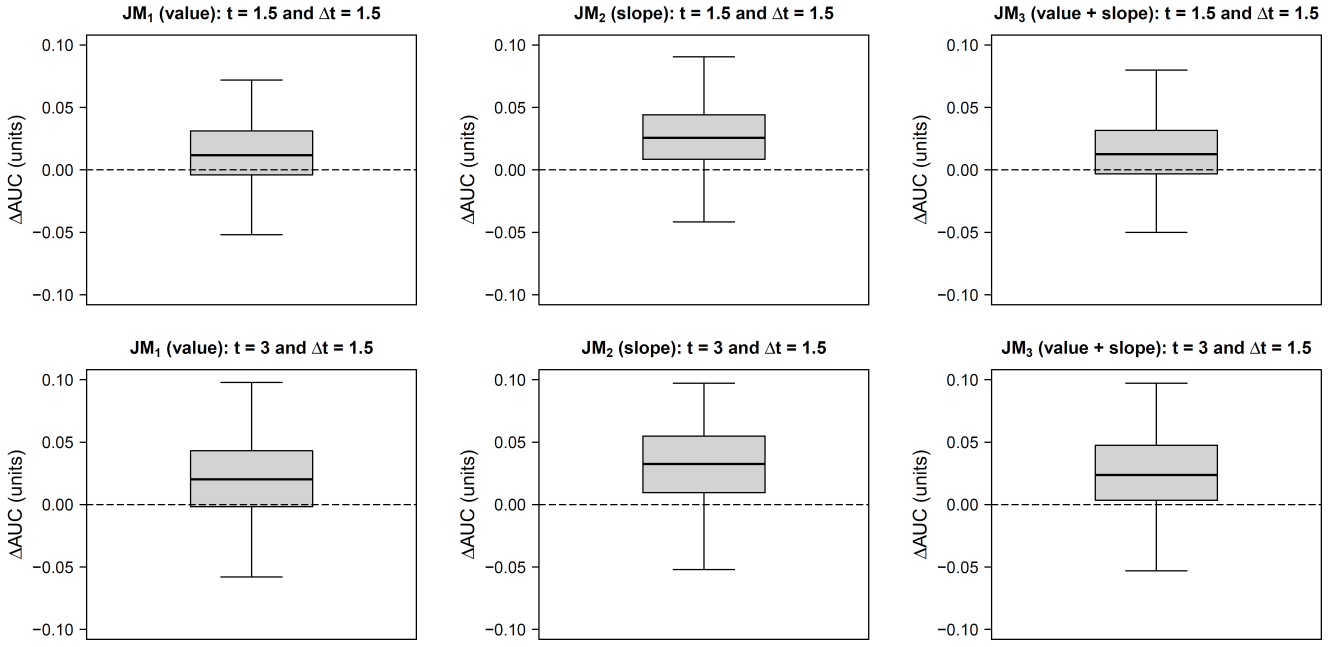

FIGURE 2 Results of a 4-fold cross-validation procedure to compare the discriminative performance of JM_1, JM_2, and JM_3 at $t \in \{1.5, 3\}$ years with $\Delta t = 1.5$. Each boxplot shows, for a given model, the $\Delta\text{AUC}(t, u)$ between a quadratic and a linear accommodations of the rescaled nadir CD4.

To conclude, Appendix D provides the mean estimates, the standard errors, and the 95% confidence intervals of the parameters when JM_2 is fitted to the HIV data, while considering both a linear and a quadratic accommodation of the rescaled nadir CD4 (Tables S15 and S16, respectively).

7 | DISCUSSION

We have analyzed the performance of an SP joint model when a fixed continuous covariate is not linearly related to the conditional log-hazard. This departure from linearity may imperil the validity of parameter estimates, so another functional relationship is needed to account for the covariate effect. Among the existing alternatives to approximate an unknown shape, the *r*th order orthogonal series expansion emerges as a useful nonparametric estimator. In our particular framework, we have proposed its use under the data-driven MAIC-based testing criterion. This combination has a twofold purpose: testing the initially hypothesized parametric linear relationship and, if it does not hold, simultaneously suggesting a specific order $r_C^* \geq 1$ to establish the truncation point for the series expansion assumed in the alternative hypothesis. An extensive numerical simulation study has been performed to assess the significance level and power of the MAIC-based test in the joint modeling setting, corroborating that the asymptotic properties of the test are retained.

As another point of special interest, we have also conducted simulations to check the model's performance in terms of discriminative ability, through the AUC value. The incorporation of a nonlinear relationship has led to gains of around 5% in the predictive accuracy. While modest, this improvement deserves recognition, given that the reason for not achieving dramatic changes may lie in the fact that we have only considered a single misspecified covariate. This is important because the maximum likelihood estimates of joint model parameters are mainly governed by the time-dependent biomarker as the number of observations per subject increases.[10] As a further research topic, it would be interesting to explore the multivariate generalization of the results presented here. However, multidimensionality usually involves an intensive computational burden, especially in the scope of joint modeling techniques.

An example of application of our research has been provided by delving deeper into the well-recognized prognostic role played by the nadir CD4 cell count among HIV-infected subjects. In the data at hand, this covariate effect has been found to be reasonably incorporated into the suggested SP joint model under a quadratic shape, leading to slight improvements in predicting the time to achieve immunity. Interestingly, the testing procedure favors the more parsimonious model based on the MAIC criterion, but such a model does not necessarily provide the most accurate fit among the considered range of orders. This is because all orders are not equally inclined to reject the null model, which is one possible criticism of the testing procedure. In this connection, exploring alternative model selection methods would be worthwhile.

ACKNOWLEDGMENTS

This research has been partially funded by the grant.

Conflict of interest

The authors declare no potential conflict of interests.

SUPPORTING INFORMATION

Supplementary information for this article is available online. Firstly, we provide a document with all relevant results concerning the distinct appendices referenced throughout the article. Secondly, we provide the R code to assess the performance of the MAIC-based test under some of the simulation scenarios considered in Section 5. The code, which takes advantage of the R package JM[27], focuses on the joint model JM_S, with 25% censoring, and is divided into the three following scripts: 1_data_splines_25.R implements the algorithm to generate the datasets subsequently fitted by JM_S, 2_typeI_pol_splines_25.R contains the steps to assess the level of robustness of the test if H_0 may incorporate additional power terms, and 3_power_pol_pol_splines_25.R includes the steps to assess the power of the test when the partially observed covariate is truly embedded in a quadratic polynomial function, while extending H_0 to incorporate additional power terms. Lastly,

we provide the HIV data that has been used to exemplify our procedure in analysis_hiv_data.RData, along with the script analysis_hiv_data.R to fit the joint model JM_2, which provides the strongest evidence against the linear relationship.

References

1. Hart JD. *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer-Verlag, New York (USA); 1997.
2. O'Sullivan F. Fast computation of fully automated log-density and log-hazard estimators. *SIAM J Sci Statist Comput*. 1988; 9(2): 363–379.
3. Ledwina T. Data-Driven Version of Neyman's Smooth Test of Fit. *J Am Stat Assoc*. 1994; 89(427): 1000–1005.
4. Kooperberg C, Stone CJ, Truong YK. Hazard Regression. *J Am Stat Assoc*. 1995; 90(429): 78–94.
5. Jain AK, Strawderman RL. Flexible hazard regression modeling for medical cost data. *Biostatistics*. 2002; 3(1): 101–118.
6. Fan J, Runze L. Variable Selection for Cox's Proportional Hazards Model and Frailty Model. *Biostatistics*. 2002; 30(1) : 74–99.
7. González-Manteiga W, Martínez-Miranda D, Van Keilegom I. Goodness-of-fit tests in proportional hazards models with random effects. *Biom J*. 2022; 65(1): 1–22.
8. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time Data. *Biostatistics*. 2000; 1(4): 465–480.
9. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: An overview. *Stat Sinica*. 2004; 14(4): 809–834.
10. Rizopoulos D, verbeke G, Molenberghs G. Shared parameter models under random effects misspecification. *Biometrika*. 2008; 95(1) : 63–74.
11. Barrett J, Diggle P, Henderson R, Taylor-Robinson D. Joint modelling of repeated measurements and time-to-event outcomes: flexible model specification and exact likelihood inference. *J Royal Stat Soc B*. 2015; 77(1): 131–148.
12. Barry D, Hartigan JA. An omnibus test for departures from constant mean. *Ann Stat*. 1990; 18(3): 1340–1357.
13. Eubank RL, Spiegelman CH. Testing Goodness of Fit of a Linear Model via Nonparametric Regression Techniques. *J Am Stat Assoc*. 1990; 85(410): 387–392.
14. Eubank RL, Hart JD. Testing Goodness-of-Fit in Regression Via Order Selection Criteria. *Ann Stat*. 1992; 20(3): 1412–1425.
15. Akaike H. A New Look at the Statistical Model Identification. *IEEE T Automat Contr*. 1974; 19(6): 716–723.
16. Aerts M, Claeskens G, Hart JD. Testing the Fit of a Parametric Function. *J Am Stat Assoc*. 1999; 94(447): 869–879.
17. Nysen R, Aerts M, Faes C. Testing goodness of fit of parametric models for censored data. *Stat Med*. 2012; 31(21): 2374–2385.
18. Baayen C, Hougaard P, Pipper CB. Testing Effect of a Drug Using Multiple Nested Models for the Dose-Response. *Biometrics*. 2015; 71(2): 417–427.
19. Sempa JB, Theresa M. Rossouw TM, Lesaffre E, Nieuwoudt M. Cumulative viral load as a predictor of CD4+ Tcell response to antiretroviral therapy using Bayesian statistical models. *PLoS One*. 2019; 14(11):e0224723 : 1–19.
20. Satten GA, Datta S, Williamson JM. Inference Based on Imputed Failure Times for the Proportional Hazards Model With Interval-Censored Data. *J Am Stat Assoc*. 1998; 93(441): 318–327.

21. Kelly C, Gaskell KM, Richardson M, Klein N, Garner P, MacPherson P. Discordant Immune Response with Antiretroviral Therapy in HIV-1: A Systematic Review of Clinical Outcomes. *PLoS One*. 2016; 11(6):e0156099 : 1–15.

22. Kullback S. *Information Theory and Statistics*. Wiley & Sons, New York (USA); 1959.

23. Woodrofe M. On Model Selection and the Arc Sine Laws. *Ann Stat*. 1982; 10(4): 1182–1194.

24. Schwartz G. Estimating the Dimension of a Model. *Ann Stat*. 1978; 6(2): 461–464.

25. Spitzer F. A combinatorial lemma and its applications to probability theory. *Trans Amer Math Soc*. 1956; 82(2): 323–339.

26. Zhang D, Davidian M. Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*. 2001; 57(3): 795–802.

27. Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data with Applications in R*. Chapman & Hall/CRC Biostatistic Series: Boca Raton, Florida (US); 2012.

28. Andrinopoulou ER, Eilers PHC, Takkenberg J, Rizopoulos D. Improved Dynamic Predictions from Joint Models of Longitudinal and Survival Data with Time-Varying Effects Using P-Splines. *Biometrics*. 2018; 74(2): 685–693.

29. Wang Y, Taylor JMG. Jointly Modeling Longitudinal and Event Time Data with Application to Acquired Immunodeficiency Syndrome. *J Am Stat Assoc*. 2001; 96(455): 895–905.

30. Guo X, Carlin BP. Separate and Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Packages. *The American Statistician*. 2004; 58(1): 16–24.

31. Aldrete S, Jang J-H, Easley KA, Okulicz J, Dai T, Chen Y-N, Pino M, Agan BK, Maves RC, Paiardini M, Marconi VC. CD4 rate of increase is preferred to CD4 threshold for predicting outcomes among virologically suppressed HIV-infected adults on antiretroviral therapy. *PLoS One*. 2020; 15(1): 1–16.