



Flexible modeling of longitudinal health-related quality of life data accounting for informative dropout in a cancer clinical trial

Audrey Winter^{1,2} · Benjamin Cuer³ · Thierry Conroy^{4,5} · Beata Juzyna⁶ · Sophie Gourgou^{1,2} · Caroline Mollevi^{1,2,3} · Célia Touraine^{1,2,3}

Accepted: 1 September 2022 / Published online: 17 September 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract

Purpose A joint modeling approach is recommended for analysis of longitudinal health-related quality of life (HRQoL) data in the presence of potentially informative dropouts. However, the linear mixed model modeling the longitudinal HRQoL outcome in a joint model often assumes a linear trajectory over time, an oversimplification that can lead to incorrect results. Our aim was to demonstrate that a more flexible model gives more reliable and complete results without complicating their interpretation.

Methods Five dimensions of HRQoL in patients with esophageal cancer from the randomized clinical trial PRODIGE 5/ ACCORD 17 were analyzed. Joint models assuming linear or spline-based HRQoL trajectories were applied and compared in terms of interpretation of results, graphical representation, and goodness of fit.

Results Spline-based models allowed arm-by-time interaction effects to be highlighted and led to a more precise and consistent representation of the HRQoL over time; this was supported by the martingale residuals and the Akaike information criterion.

Conclusion Linear relationships between continuous outcomes (such as HRQoL scores) and time are usually the default choice. However, the functional form turns out to be important by affecting both the validity of the model and the statistical significance.

Trial registration This study is registered with ClinicalTrials.gov, number NCT00861094.

Keywords Joint model · Linear mixed model · Spline · Health-related quality of life · Clinical trials · Cancer

✉ Audrey Winter
audrey.winter@icm.unicancer.fr

Benjamin Cuer
cuerbenjamin@gmail.com

Thierry Conroy
t.conroy@nancy.unicancer.fr

Beata Juzyna
b-juzyna@unicancer.fr

Sophie Gourgou
sophie.gourgou@icm.unicancer.fr

Caroline Mollevi
caroline.mollevi@icm.unicancer.fr

Célia Touraine
celia.touraine@icm.unicancer.fr

¹ Biometrics Unit, Cancer Institute of Montpellier, University of Montpellier, Parc Euromédecine, 208 Avenue des Apothicaires, 34090 Montpellier, France

² French National Platform Quality of Life and Cancer, Montpellier, France

³ Desbrest Institute of Epidemiology and Public Health, University of Montpellier, INSERM, 641 Avenue du Doyen Gaston Giraud, Montpellier, France

⁴ Department of Medical Oncology, Institut de Cancérologie de Lorraine, 6 Av. de Bourgogne, Vandoeuvre-lès-Nancy, France

⁵ Teams MICS, APEMAC, Université de Lorraine, Nancy, France

⁶ R&D Unicancer, Paris, France

Introduction

Time-to-event outcomes have often been the primary endpoint in clinical trials, with health-related quality of life (HRQoL) relegated to a secondary focus. However it is now well known that attention must be paid to both “quantity” and quality of life. Generally, HRQoL data are collected during trials through standardized questionnaires completed by patients at selected time points of clinical interest (e.g., baseline, and before and after treatment).

HRQoL and time-to-event analyses are usually done separately using linear mixed models (LMMs) and survival models, respectively. This kind of approach may however lead to inefficient or biased results [1–3]. In particular, when studying the HRQoL longitudinal outcome, dropouts caused by events or simply nonresponse might be observed. When the dropout mechanism depends on unobserved longitudinal HRQoL measurements, analysis based on an LMM alone can be misleading [2, 4]. Thus, to overcome this problem, it is necessary to jointly model the HRQoL outcome and the risk of dropout through a joint modeling framework. A joint model (JM) is composed of two submodels: a model for the time-to-event outcome (e.g., a proportional hazard model) and a model for the longitudinal outcome (e.g., an LMM) linked together though shared random effects.

Furthermore, the LMM used to model the HRQoL outcome is usually a random intercept and slope model, assuming a linear trajectory over time. This oversimplification can lead to wrong or simplistic findings. As an example, a linear trajectory will not be able to capture nonmonotonic evolution, typically, a deterioration during the treatment phase followed by the patient’s health improvement.

Currently, the analysis of HRQoL in clinical trials is often brief and simplistic and ignores the occurrence of dropouts. In a previous article [5], a simulation study demonstrated the benefit of using a JM rather than an LMM alone when death might interrupt observation of the HRQoL outcome, resulting in an informative dropout. The two models compared assumed linear trajectories for the HRQoL outcome. However, this assumption could not hold on actual data, resulting in unsatisfactory results. In this article, through the application of the models to data from a clinical trial of patients with advanced esophageal cancer, the aim is to compare two JMs in terms of interpretation of the results, graphical representation, and goodness of fit: a JM assuming a linear trajectory of the HRQoL outcome and a spline-based JM allowing for a more flexible trajectory.

In this article, we first present the clinical trial, analyze completion of the questionnaires to explain why a

JM should be used rather than an LMM, and study the trajectories of the HRQoL outcome to highlight the fact that a random intercept and slope LMM might not be the best option. Second, we detail the two models and their assumptions; we also detail how to make a choice between the two. Finally, we discuss our findings.

Motivating dataset

The clinical trial PRODIGE5/ACCORD17

PRODIGE 5/ACCORD 17 (NCT00861094) was a multicenter, randomized, open-label, parallel-group, phase 2–3 clinical trial comparing two treatment regimens of definitive chemoradiotherapy: FOLFOX (the combination of oxaliplatin and fluorouracil with leucovorin) versus fluorouracil and cisplatin. A total of 267 patients (134 in the experimental arm, and 133 in the control arm) with esophageal cancer were enrolled between October 2004 and August 2011 [6, 7]. Progression-free survival (PFS) was the primary endpoint of this study, and no significant difference was highlighted between the two treatment groups. Overall survival (OS) and HRQoL were also included as secondary endpoints. Just as for PFS, no significant difference in OS was found between the groups. HRQoL was assessed by means of the European Organisation for Research and Treatment of Cancer core quality-of-life questionnaire (QLQ-C30 version 3.0 composed of 30 items grouped into five functional scales, nine symptoms scales and a global health status/HRQoL scale) [8] and the esophagus-specific questionnaire (QLQ-OES18 composed of 18 items for ten symptoms scales) [9] at inclusion, during treatment (at 1.25 and 3 months), at the first evaluation of treatment efficacy (month 4), and during follow-up (at 6, 12, 24, and 36 months). In general, a high score for a functional scale and the global health status/HRQoL scale represents a high level of functioning/HRQoL while, for symptom scales, it represents a high level of symptomatology. The secondary objective was mostly based on QLQ-C30 scales: of global health status/HRQoL (QL), physical functioning (PF), fatigue (FA), and pain (PA) but also included dysphagia (OESDYS) from the QLQ-OES18 scale. Hereafter, all analyses are based only on these dimensions. Of note, unlike the symptom scales of the QLQ-C30 (i.e., FA and PA), a high OESDYS score represents low symptoms.

HRQoL questionnaire completion

At each assessment, several scenarios of completion are possible; a questionnaire can be fully completed, partially completed or entirely noncompleted. Accordingly, depending on which items have been completed or not, one score might

be missing and another might not. Hereafter, all analyses are performed separately for each scale, making no distinction between missing score data (i.e., whether due to fully noncompleted questionnaire or item non-response). Barplots in Fig. 1 show the percentages of completion at each visit for each of the five scales of interest. The same conclusions can be drawn for both arms, namely a decrease in completion over time (in yellow in Fig. 1). Noncompletion across visits gives rise to two types of missing data. In the first, patients do not fill in some items or the entire questionnaire at particular visits and then fill it in again at later times; this corresponds to intermittent missing data (in blue in Fig. 1), assumed to be missing at random in our analysis. In the second type, patients do not fill in some items or the entire questionnaire at a particular visit and then do not complete any further items/questionnaires; this corresponds to monotone missing data (in gray in Fig. 1), considered to be missing not at random in our analysis. Hereafter, we use the term dropout to refer to monotone missing data, whether there is an actual

dropout or simply nonresponse without any associated event being identified. Those dropouts, which can be informative (i.e., linked to the HRQoL), have to be taken into account to correctly model the HRQoL, hence the importance of a joint modeling approach as outlined previously.

HRQoL over time

For each dimension studied, individual trajectories by arm were plotted (Fig. 2). LOESS (locally estimated scatterplot smoothing) curves were also drawn by treatment groups to get a better idea of the mean trajectories of the observed scores. The same trend was observed for every dimension, namely an HRQoL deterioration (i.e., decrease in QL, PF, and OESDYS, increase in PA and FA) from baseline ($t=0$) to the final day of radiotherapy or on day 1 of the last chemotherapy cycle ($t=1.25$ months/3 months), then an improvement (i.e., increase in QL, PF, and OESDYS, decrease in PA and FA) up to 6 months, and finally a plateau until the

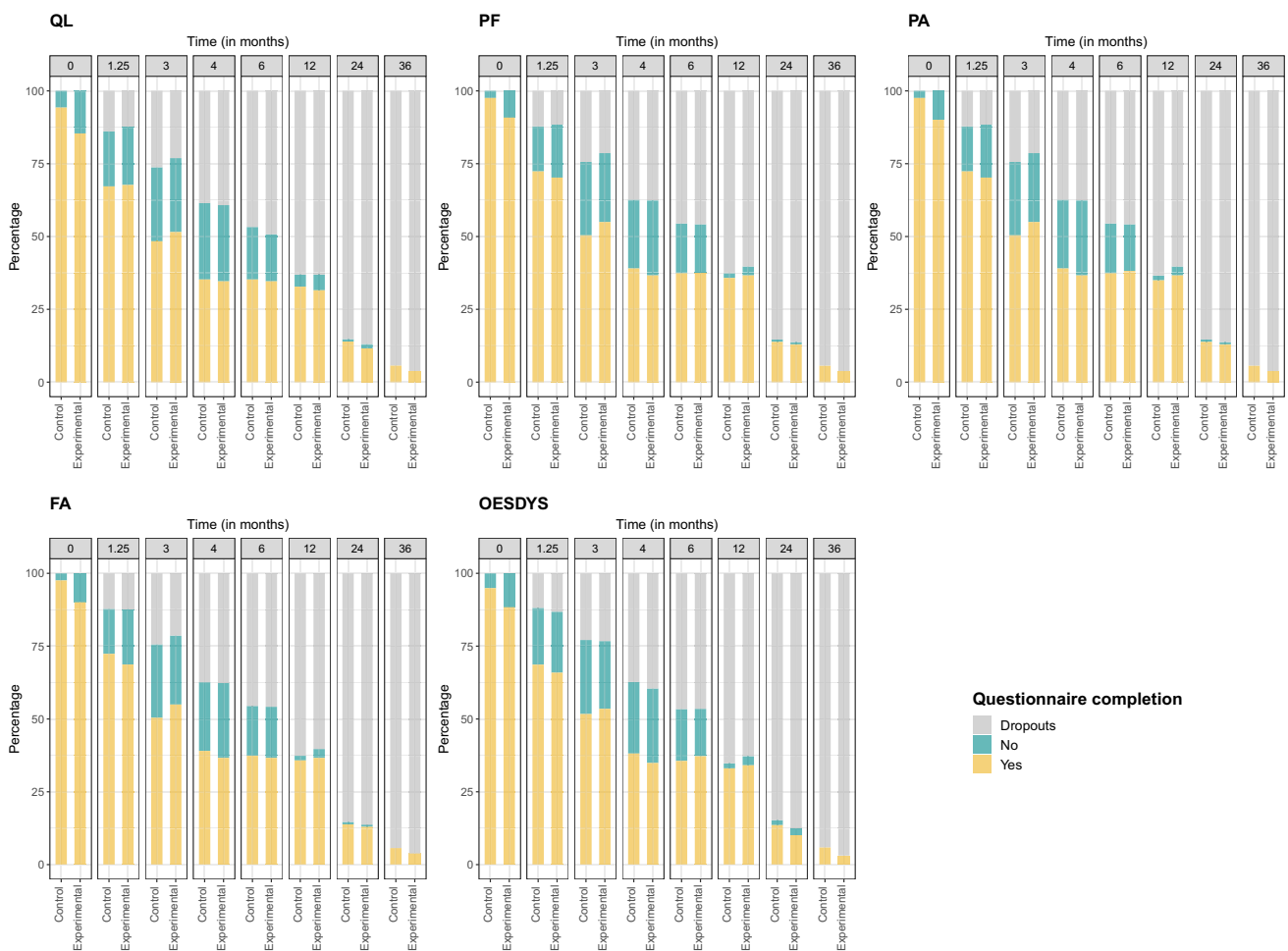


Fig. 1 QLQ-C30 questionnaire completion over time by arm on global health status/HRQoL (QL), physical functioning (PF), fatigue (FA), pain (PA), and dysphagia (OESDYS) scales. (Color figure online)

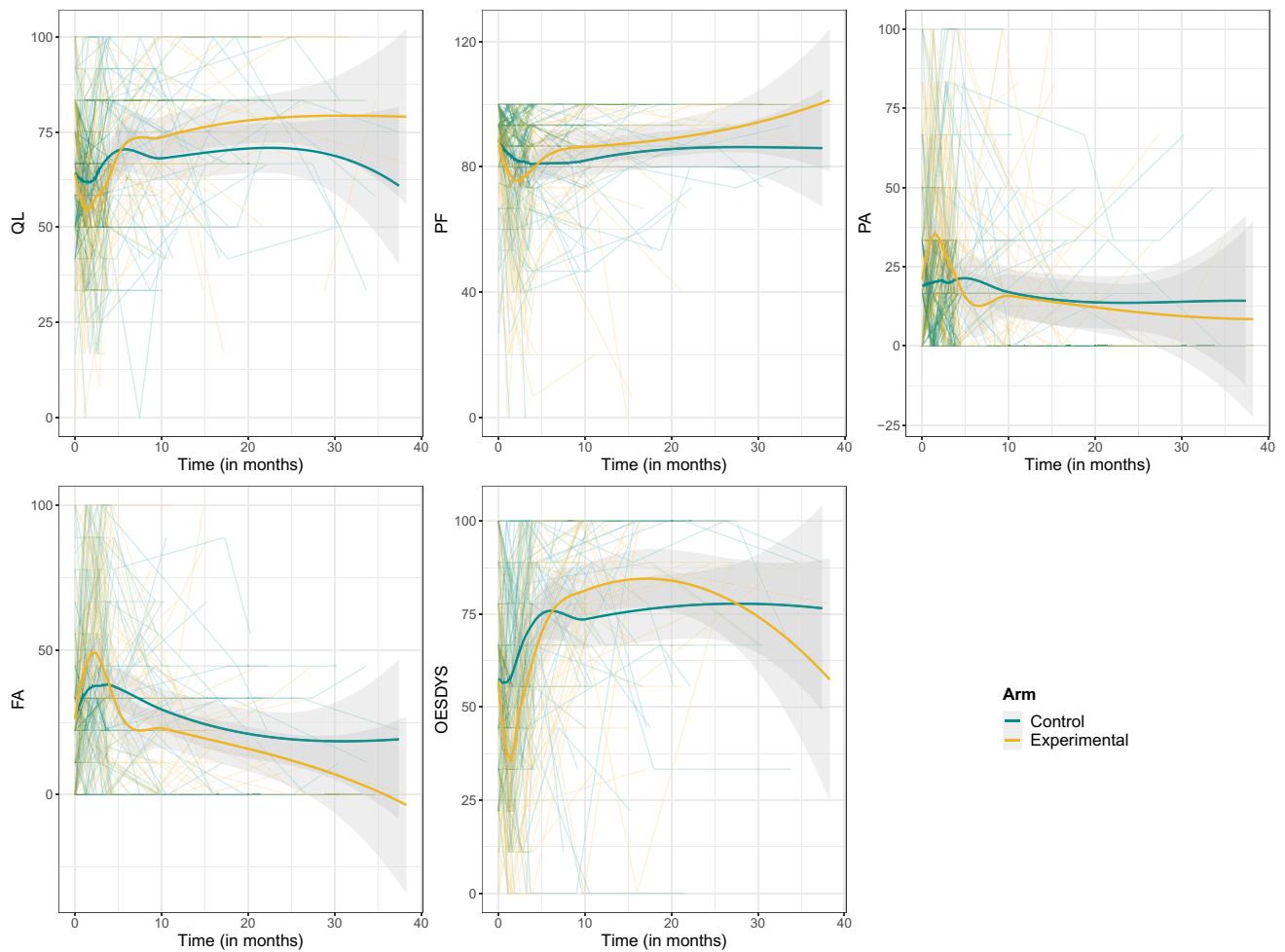


Fig. 2 Individual trajectories and LOESS (locally estimated scatterplot smoothing) regression smoothing over time by arm on global health status/HRQoL (QL), physical functioning (PF), fatigue (FA), pain (PA), and dysphagia (OESDYS)

end. Therefore, HRQoL trajectories, whatever the dimension considered, do not seem to follow a linear trend; it therefore appears to be relevant to consider a more flexible modeling approach allowing nonlinear trajectories.

Joint model

As mentioned previously, generally a JM is composed of two submodels: an LMM for the longitudinal outcome and a survival model. In the two considered JMs, the general form of the HRQoL score of patient i at time t , $Y_i(t)$, is:

$$\begin{cases} Y_i(t) = Y_i^*(t) + \varepsilon_i(t) \\ Y_i^*(t) = \beta^T X_i(t) + b_i^T Z_i(t) \\ b_i \sim N(0, D), \varepsilon_i(t) \sim N(0, \sigma^2) \end{cases} \quad (1)$$

where $X_i(t)$ and $Z_i(t)$ are the respective time-dependent design matrices for the fixed and random effects, β and b_i , containing the covariates values for each subject (rows representing patients and columns variables), and $\varepsilon_i(t)$ is the error terms, also time-dependent. Roughly the fixed effects have the same interpretation as in a linear regression and characterize the mean score trajectory over time [i.e., $\mathbb{E}(Y_i(t)) = \beta^T X_i(t)$], and the random effects b_i describe the deviations for each patient i from the mean trajectory, therefore, combining with the fixed effects, depict the individual trajectories. We assume that b_i and $\varepsilon_i(t)$ are independent and normally distributed with mean zero and variance–covariance matrix/variance D and σ^2 , respectively. Each parameter is defined in more details thereafter in Sections “[Linear HRQoL trajectory](#)” and “[Spline-based HRQoL trajectory](#)”, according to each LMM. Finally, $Y_i^*(t)$ denotes the true value of the longitudinal outcome at time point t .

This term was included in a proportional hazards model used for the time-to-dropout:

$$h_i(t) = h_0(t)\exp\{\gamma arm_i + \alpha Y_i^*(t)\} \tag{2}$$

where $h_0(t)$ represents the baseline hazard function (describing the instantaneous risk of dropout all covariates being equal to zero), arm_i the arm factor equals 1 if patient i belongs to the experimental arm and 0 if patient i belongs to the control arm, γ its corresponding effect and α the parameter of association that quantifies the effect of the current HRQoL value on the risk of dropout.

Linear HRQoL trajectory

In the first model, we assume that the HRQoL score follows a linear trajectory over time using a random coefficient model, as is commonly done in clinical trial settings:

$$\begin{cases} Y_i^*(t) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t + \beta_2(arm_i \times t) \\ \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N(0, D), D = \begin{pmatrix} \sigma_{b_0}^2 & \sigma_{b_0 b_1} \\ \sigma_{b_0 b_1} & \sigma_{b_1}^2 \end{pmatrix} \end{cases} \tag{3}$$

where the fixed intercept β_0 represents the mean score at inclusion ($t = 0$), the fixed slope β_1 the score change by unit of time in the control arm, the interaction effect β_2 the difference between the slopes of the experimental and control arms ($\beta_1 + \beta_2$ being the slope in the experimental arm), and the random intercept b_{0i} and random slope b_{1i} the individual deviations from the fixed intercept and fixed slope, respectively. Of note, no arm effect was added to the model, as HRQoL at baseline was broadly similar between the two arms due to randomization.

Spline-based HRQoL trajectory

In the second model, we suppose that the HRQoL trajectory is not linear over time, since, as can be seen in Fig. 2, this assumption seems to be wrong. To gain flexibility, a model using natural splines was investigated:

$$\begin{cases} Y_i^*(t) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})B_n(t, \lambda_1) + (\beta_2 + b_{2i})B_n(t, \lambda_2) + (\beta_3 + b_{3i})B_n(t, \lambda_3) \\ \quad + (\beta_4 + b_{4i})B_n(t, \lambda_4) + \beta_5\{B_n(t, \lambda_1)arm_i\} + \beta_6\{B_n(t, \lambda_2)arm_i\} \\ \quad + \beta_7\{B_n(t, \lambda_3)arm_i\} + \beta_8\{B_n(t, \lambda_4)arm_i\} \\ \begin{pmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \\ b_{3i} \\ b_{4i} \end{pmatrix} \sim N(0, D), \text{ with } D \text{ unstructured} \end{cases} \tag{4}$$

where $\{B_n(t, \lambda_k); k = 1, 2, 3, 4\}$ represents a B-spline basis matrix for a natural cubic spline with three internal knots placed at 1.25, 3, and 6 months. Based on the observations made in Section “HRQoL questionnaire completion”, these time points correspond to observed changes in trend. The fixed intercept β_0 still represents the mean score at inclusion, the fixed effects $\beta_1, \beta_2, \beta_3,$ and β_4 govern the score trajectory in the control arm, the fixed effects $\beta_5, \beta_6, \beta_7,$ and β_8 allow the trajectory in the experimental arm to be different from the one in the control arm, and the random effects $b_{0i}, b_{1i}, b_{2i}, b_{3i},$ and b_{4i} allow for individual deviations from the mean trajectories.

A spline function is a piecewise polynomial function, the time variable being divided into distinct intervals (using the knots) where pieces join smoothly. Therefore, its general formulation is not a linear combination of $\{\text{intercept}, t, t^2, \dots\}$, as in a usual polynomial, but a set of functions depending on time and these intervals. As a consequence, spline coefficients β_{1-8} cannot be interpreted directly, in contrast to those from the linear model, in which it is simply a linear combination of the intercept and t in each arm. In particular, a more restrictive model, usually used for inference purpose, is much more interpretable than a flexible model for which prediction is the aim; this representing the tradeoff between flexibility and interpretability [10]. Therefore, for this analysis, more importance has to be given to the predicted trajectories for instance.

Arm-by-time interaction effect on HRQoL

Evaluating the arm-by-time effect in the linear trajectory JM is a straightforward task; attention has simply to be paid to the interaction term coefficient estimate (i.e., $\hat{\beta}_2$ in Eq. 3). However, for the spline-based JM, the interpretation of the spline coefficients is not that simple; however the arm-by-time interaction effect can be tested using a log-likelihood ratio (LR) test with $H_0 : \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$ vs $H_1 : \beta_5 \neq 0$ or $\beta_6 \neq 0$ or $\beta_7 \neq 0$ or $\beta_8 \neq 0$ (cf Eq. 4).

Joint model assumptions

For the survival submodel, martingale residuals plots were used to check the excess observed events and the chosen functional form (identity in our case) for the time-dependent covariate (current true score value in our case) and the Cox-Snell residuals to assess about the overall goodness-of-fit of the submodel [4, 11]. For the longitudinal submodel, two types of widely used residuals in mixed models were explored; subject-specific (conditional) residuals for checking the homoscedasticity and normality assumptions, and marginal (population averaged) residuals to investigate misspecification of the mean structure $\beta^T X_i(t)$ and validate the assumptions for the within-subjects covariance structure [4, 12, 13]. However, a problem arises when using these residuals in the context of joint modeling. Indeed, scores might be missing from a given time point (i.e., patients might drop-out) which, in turn, affects the use of residuals based on the observed data alone [4, 14, 15]. To overcome this nonrandom dropout issue, a multiple imputation approach was performed (10 imputations) to obtain scores that would have been observed if patients had not dropped out of the study. The residuals produced from the augmented longitudinal data were then used to check the JM assumptions by means of the usual diagnostic plots. More details about JM assumptions diagnostic can be found in Supplementary materials—Section 1.

Application

All analyses were performed in a modified intent-to-treat population; all patients with at least one available HRQoL score were included. Thus, the population of analysis was not the same across the five dimensions explored: QL (experimental arm: $N=130$; control arm: $N=122$), PF (experimental arm: $N=131$; control arm: $N=123$), PA (experimental arm: $N=131$; control arm: $N=123$), FA (experimental arm: $N=131$; control arm: $N=123$), and OESDYS (experimental arm: $N=129$; control arm: $N=118$). The event of interest considered in the survival submodel of the JM was the dropout defined as the last visit for which a score is available (right-censored time in case of an available score at the last planned visit), and the follow-up ended with the occurrence of one of these events.

All analyses were performed with R version 4.0.3 [16] using the *JM* package [17] for the joint modeling approach and the *splines* package for the spline-based modeling strategy. R codes are available from the first author under reasonable request. Main models implementation can be found in Supplementary materials—Section 2, more details in [4].

Comparison of linear and spline-based joint models

JMs for linear and spline-based HRQoL trajectories were applied to the five scales: QL, PF, PA, FA, and OESDYS. The logarithm of the baseline hazard function was approximated by B-splines, using four knots placed at the quantiles of the event times. For each JM, seven Gauss-Hermite pseudo-adaptive quadrature points were used to approximate the integrals over the random effects. Tables 1 and 2 give an overview of the estimates obtained from these two strategies.

Concerning the survival submodel, no arm effect γ was identified whatever the model and/or the dimension, meaning that the risk of dropout did not differ from one arm to the other, which is consistent with Fig. 1. The association effect between the longitudinal outcome and the risk of dropout α was significant for PF and FA in the linear JM ($\hat{\alpha}_{PF} = -0.020 [-0.031; -0.010]$, $p < 0.001$ and $\hat{\alpha}_{FA} = 0.013 [0.004; 0.023]$, $p = 0.005$); this was confirmed in the spline-based JM ($\hat{\alpha}_{PF} = -0.017 [-0.025; -0.009]$, $p < 0.001$ and $\hat{\alpha}_{FA} = 0.012 [0.005; 0.018]$, $p < 0.001$)—this means a decrease in PF score/increase in FA score corresponded to an $\exp(-\hat{\alpha})$ -fold increase in the risk of dropout. For PA and OESDYS, no association effect was significant either in the linear or in the spline-based JM. However, for QL, no significant association effect was highlighted in the linear model ($\hat{\alpha}_{QL} = -0.011 [-0.027; 0.005]$, $p = 0.162$), while in the spline-based model a decrease in QL score was found to be associated with an increased risk of dropout ($\hat{\alpha}_{QL} = -0.027 [-0.041; -0.014]$, $p < 0.001$).

Concerning the longitudinal outcome submodel, coefficient estimates cannot be compared directly (except $\hat{\beta}_0$), since the interpretation of the regression coefficients is different across models. Moreover, as noted previously in Section “Spline-based HRQoL trajectory”, spline coefficients β_{1-8} cannot be interpreted directly, that’s the reason why they are not presented in this article. Thus, the comparison was mainly based on a graphical representation of the predicted mean HRQoL trajectories conditionally to the arm (i.e., $\mathbb{E}(Y_i(t) | arm_i)$, Fig. 3), which illustrates the goodness of fit of each model considering the mean trajectories obtained from the raw data (Fig. 2). Visually, linear trajectories showed an increase in QL and OESDYS and a decrease in PF, PA, and FA scores over time; trajectories by arm being equal at baseline without any possible intersect thereafter. Regarding the arm-by-time interaction effect, none was found whatever the dimension. With nonlinear trajectories, the trend was the same as that shown in Fig. 2. First, a deleterious effect was observed in both arms (stronger in the experimental arm, FOLFOX toxicities are known to be considerable), namely a decrease in PF, QL, and OESDYS as well as an increase in PA and FA scores between baseline and month 1.25 or 3 (i.e., on the final day of radiotherapy or on day 1 of the last chemotherapy cycle). Then,

Table 1 Parameter estimates with 95% confidence intervals (95% CI) of the linear joint model applied to PRODIGES5/ACCORD17 HRQoL data in five dimensions: global health status/HRQoL (QL), physical functioning (PF), fatigue (FA), pain (PA), and dysphagia (OESDYS)

Dimension			Estimate [95% CI]	<i>p</i> value	
QL	Longitudinal sub model	β_0	61.734 [59.711; 63.758]	<0.001	
		β_1	0.325 [0.039; 0.611]	0.026	
		β_2	0.179 [− 0.195; 0.553]	0.348	
	Time-to-event sub model	γ	0.006 [− 0.250; 0.262]	0.963	
		α	− 0.011 [− 0.027; 0.005]	0.162	
		AIC	8569.235		
		BIC	8629.235		
	PF	Longitudinal sub model	β_0	83.899 [81.897; 85.902]	<0.001
			β_1	− 0.520 [− 0.933; − 0.107]	0.014
			β_2	0.069 [− 0.416; 0.553]	0.781
Time-to-event sub model		γ	− 0.043 [− 0.299; 0.212]	0.740	
		α	− 0.020 [− 0.031; − 0.010]	<0.001	
		AIC	8977.056		
		BIC	9037.191		
PA		Longitudinal sub model	β_0	23.822 [21.220; 26.425]	<0.001
			β_1	− 0.212 [− 0.586; 0.163]	0.268
			β_2	− 0.211 [− 0.692; 0.270]	0.390
	Time-to-event sub model	γ	− 0.060 [− 0.316; 0.196]	0.645	
		α	0.008 [− 0.004; 0.020]	0.196	
		AIC	9647.340		
		BIC	9707.475		
	FA	Longitudinal sub model	β_0	35.955 [33.142; 38.767]	<0.001
			β_1	− 0.213 [− 0.565; 0.139]	0.236
			β_2	− 0.227 [− 0.691; 0.238]	0.339
Time-to-event sub model		γ	− 0.035 [− 0.290; 0.221]	0.791	
		α	0.013 [0.004; 0.023]	0.005	
		AIC	9601.633		
		BIC	9661.768		
OESDYS		Longitudinal sub model	β_0	55.444 [52.493; 58.395]	<0.001
			β_1	1.015 [0.541; 1.490]	<0.001
			β_2	0.108 [− 0.474; 0.690]	0.717
	Time-to-event sub model	γ	− 0.048 [− 0.316; 0.219]	0.724	
		α	− 0.011 [− 0.027; 0.004]	0.158	
		AIC	9132.270		
		BIC	9191.929		

AIC Akaike information criterion, BIC Bayesian information criterion

HRQoL improved and stabilized during follow-up for the five dimensions. Arm trajectories intersected around month 6, after which better HRQoL was observed in the experimental arm. Of note, confidence intervals became wider over time, making estimations less precise. This uncertainty can be explained by the discontinued collection of the HRQoL scores due to dropouts. Unlike the linear JM, in which the arm-by-time interaction term β_2 was not significant for any dimension (Table 1), the spline-based JM found an arm-by-time interaction effect for QL, PF, PA and OESDYS (LR test $p_{QL} = 0.040$, $p_{PF} = 0.017$, $p_{PA} = 0.003$ and $p_{OESDYS} < 0.001$) and a trend for FA ($p_{FA} = 0.076$) highlighting between-arms differences in HRQoL score trajectories.

Tables 1 and 2 also present Akaike information criterion (AIC) and Bayesian information criterion (BIC) values for each model. The AIC always chose the spline-based model over the linear one ($\Delta AIC_{QL} = 31.78$, $\Delta AIC_{PF} = 127.60$, $\Delta AIC_{PA} = 43.10$, $\Delta AIC_{FA} = 130.07$, and $\Delta AIC_{OESDYS} = 87.01$). The BIC, meanwhile, preferred the linear model for QL and PA ($\Delta BIC_{QL} = -31.75$ and $\Delta BIC_{PA} = -20.57$) and the spline-based model for PF, FA and OESDYS ($\Delta BIC_{PF} = 63.93$, $\Delta BIC_{FA} = 66.40$, and $\Delta BIC_{OESDYS} = 23.84$), which might be explained by a more conservative penalty in the BIC that will always tend to select simpler models (i.e., those containing fewer parameters).

Table 2 Parameter estimates with 95% confidence interval (95% CI) of the spline-based joint model applied to PRODIGE5/ACCORD17 HRQoL data in five dimensions: global health status/HRQoL (QL), physical functioning (PF), fatigue (FA), pain (PA), and dysphagia (OESDYS)

Dimension			Estimates [95% CI]	<i>p</i> value
QL	Longitudinal sub model	β_0	64.203 [61.801; 66.606]	<0.001
		γ	– 0.049 [– 0.310; 0.213]	0.714
	Time-to-event sub model	α	– 0.027 [– 0.041; – 0.014]	<0.001
		AIC	8537.459	
		BIC	8660.989	
	PF	Longitudinal sub model	β_0	89.025 [87.073; 90.976]
γ			– 0.050 [– 0.303; 0.204]	0.700
Time-to-event sub model		α	– 0.017 [– 0.025; – 0.009]	<0.001
		AIC	8849.458	
		BIC	8973.265	
PA		Longitudinal sub model	β_0	20.046 [16.765 23.327]
	γ		– 0.049 [– 0.304; 0.205]	0.703
	Time-to-event sub model	α	0.006 [– 0.003; 0.014]	0.180
		AIC	9604.237	
		BIC	9728.044	
	FA	Longitudinal sub model	β_0	26.535 [23.412; 29.659]
γ			– 0.008 [– 0.260; 0.245]	0.953
Time-to-event sub model		α	0.012 [0.005; 0.018]	<0.001
		AIC	9471.560	
		BIC	9595.367	
OESDYS		Longitudinal sub model	β_0	57.638 [54.322; 60.953]
	γ		– 0.029 [– 0.291; 0.233]	0.828
	Time-to-event sub model	α	– 0.007 [– 0.017; 0.002]	0.125
		AIC	9045.259	
		BIC	9168.087	

AIC Akaike information criterion, BIC Bayesian information criterion

Joint model assumptions

According to Figs. S1–S5 in the Supplementary materials, for the longitudinal outcome submodel a random behavior around zero was observed for the multiply imputed residuals for all dimensions regardless of the model used, indicating compliance with the assumptions mentioned in Section “Joint model assumptions”. For all Cox-Snell residual plots, the survival function estimate was around the unit exponential distribution; the models are thus considered to fit the data well. However, more attention must be paid to the martingale residuals. Across all martingale residual plots, the same trend was highlighted, namely a deviation from zero for low longitudinal responses observed for PF, PA, and FA (high responses for QL and OESDYS); that this phenomenon was much less pronounced for the spline-based JM suggests that the functional form for the time-dependent covariate (i.e., HRQoL score over time) might be more appropriate in this model.

Discussion

In this article, through the application to five HRQoL dimensions using data from the PRODIGE 5/ACCORD 17 clinical trial, we demonstrated that, despite the use of a spline-based JM that might appear more complicated than a linear JM at first sight, the interpretation of the results, is not necessarily so. On the contrary; by revealing all the information that can be extracted, such a model can facilitate the interpretation of the results and remove some of the inconsistencies that may arise from use of the usual models.

Indeed, by modeling more flexible trajectories of the longitudinal outcome, the spline-based JM, in contrast with the linear JM, allowed detection of between-arm differences in HRQoL score trajectories and association effects between the longitudinal outcome and the time-to-event. In our application, the spline-based JM identified significant arm-by-time interaction effects for almost all dimensions (a trend for FA) and a significant association effect for QL, while these effects were not significant using the linear JM. The flexible JM also gave a more reliable and precise representation of the HRQoL score trajectories, while the linear specification of the model

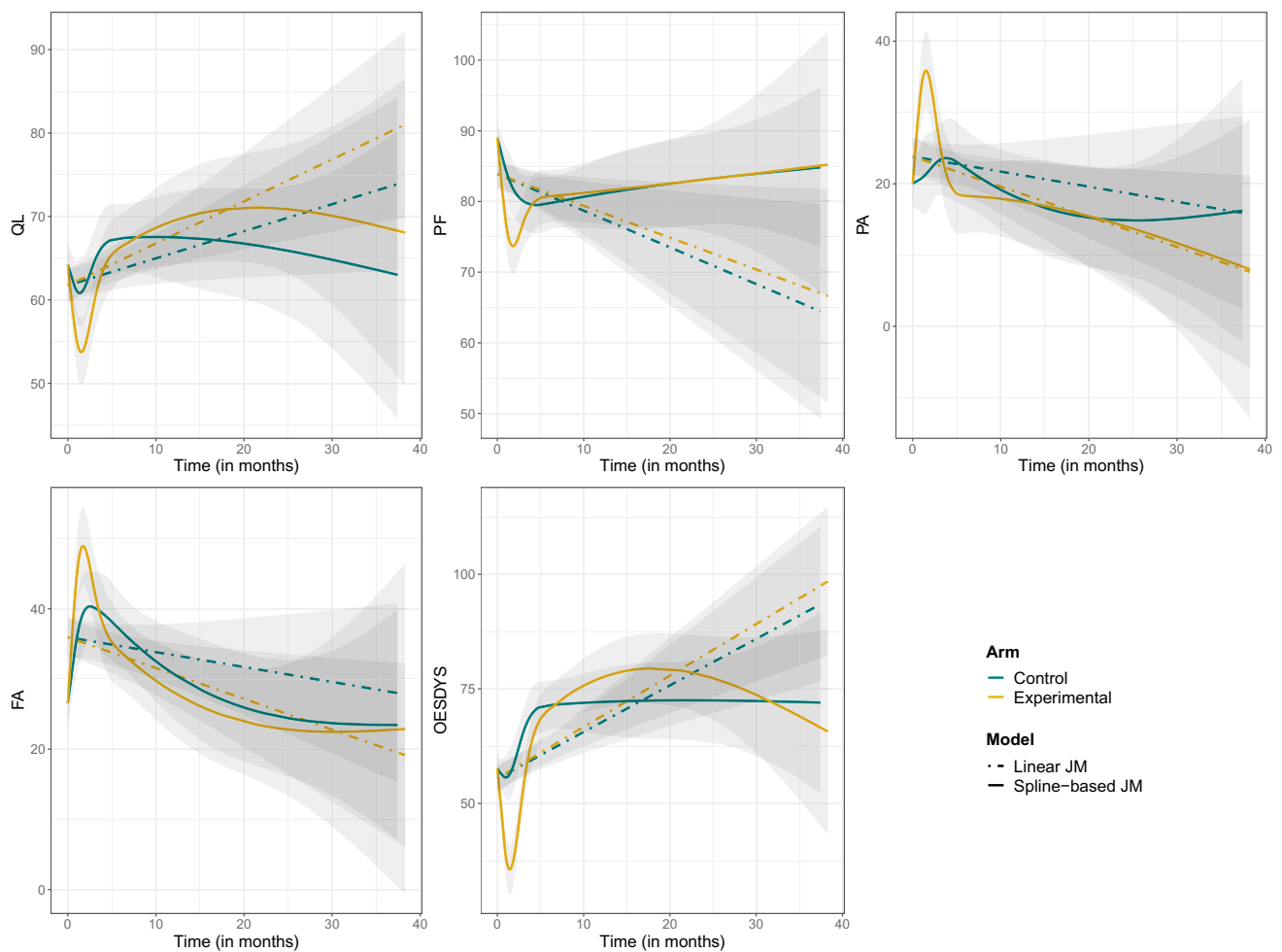


Fig. 3 Predicted mean score trajectories with confidence bands by arm for the joint models (JM) assuming linear or spline-based trajectories on global health status/HRQoL (QL), physical functioning (PF), fatigue (FA), pain (PA), and dysphagia (OESDYS)

made the score increase/decrease necessarily constant and made the beneficial effect of the experimental treatment grow throughout the whole period (the trajectories by arm being equal at baseline without any possible intersect thereafter). By contrast, the spline-based JM highlighted a nonconstant evolution of the scores and variations in the apparent benefit from one arm to the other (with trajectories intersecting once or twice). Moreover, when we checked the validity of the JM assumptions, martingale residuals suggested that the functional form of the HRQoL scores across time in the spline-based JMs might be more appropriate.

However, some aspects of applying a flexible JM might constitute limitations. For instance, in spline-based JMs, a choice had to be made concerning the number and location of the knots. Our choice was based on the trend seen in the observed mean trajectories plot but also on clinical considerations. Three knots were chosen; on the final

day of radiotherapy ($t = 1.25$ months), on day 1 of the last chemotherapy cycle ($t = 3$ months), and on the first day of follow-up ($t = 6$ months). We used the AIC to check that using fewer knots would not have been as or more efficient [18]. Thus, three knots were retained in all models, which seemed to be the best compromise between flexibility and overfitting. Of note, there were between 155 and 318 observations in the intervals defined through these knots.

Conclusion

In the literature, the choice of the functional form of a continuous variable has been widely addressed in Sauerbrei et al. articles [19, 20], Harrell's book [18] or Royston and Sauerbrei's [21]. However, from our knowledge, few articles deal with a joint modeling approach using splines in an HRQoL data context [22–24]. In Li et al. [22], a

semiparametric JM for terminal trend of HRQoL has been used, meaning that the time is counting backward from the time of death. In Yang et al. [23], the association structure between survival and longitudinal processes is of different nature and a cure model has been used for the survival part. Finally, the model in Terrin et al.'s article [24] is the closest from ours but focus is given to the survival part and not the longitudinal part. As discussed in Sauerbrei et al. [19], a linear relationship between the outcome and continuous variables (HRQoL score across time in our case) is usually the default choice, and flexible modeling techniques (i.e., using fractional polynomials, splines) are underused. However, the choice of the functional form for a variable turns out to be important, as it affects the validity of the model and the statistical significance of the variable [19]. This article illustrates this statement in a joint modeling framework but can be extrapolated to other types of models, in particular to LMMs.

In this paper, we have given a lot of arguments favoring the use of a spline-based JM to study HRQoL score trajectories in a clinical trial context. As each database is unique, here are some guidelines, with some references alongside, for helping to choose the most appropriate model:

1. Look at missing values. If a lot of monotone missing HRQoL data is to be counted, the use of a JM that uses the last visit as the time-to-event might be preferable to use of an LMM alone [5].
2. Choose the HRQoL score trajectory model carefully (whether a JM is used or not, these recommendations are still valid):
 - a. Plot the observed mean trajectory (e.g., using LOESS regression smoothing) to get an idea of the degree of flexibility needed:
 - i. If the relationship between the HRQoL outcome and time seems to be linear, a random intercept and slope model is the best choice, as it gives a better interpretability of the model parameters;
 - ii. Otherwise, a more flexible model is needed, using fractional polynomials, splines, or simply by adding, for example, a quadratic term [18, 19, 21].
 - b. Inspect the residual plots to validate the model's assumptions, and if these are not respected make the appropriate changes and/or look at the results with a critical eye [4, 21].
 - c. If hesitating between several models, consider looking at the AIC or BIC.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11136-022-03252-6>.

Acknowledgements We thank the editor for his thoughtful comments and constructive suggestions.

Funding This work was supported by the “Ligue nationale contre le cancer” (“Projet de Recherche en Epidémiologie Ligue” 2019) and the “SIRIC Montpellier Cancer” (Grant INCa_Inserm_DGOS_12553).

Data availability The dataset from the PRODIGE 5/ACCORD 17 clinical trial is not publicly available due to confidentiality requirements. Data are however available from the main coordinator, Pr. Thierry Conroy, upon reasonable request, and with permissions of the study sponsor UNICANCER R&D.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Ethical approval UNICANCER R&D, the sponsor of the PRODIGE 5/ACCORD 17 trial (ClinicalTrials.gov Identifier: NCT00861094), provided permission for the data base access.

Consent to participate Informed consent was obtained from all individual participants included in the study. Patient consent was not required for this study as we performed a secondary analysis on existing data.

References

1. Wu, L., Liu, W., Yi, G. Y., & Huang, Y. (2011). Analysis of longitudinal and survival data: Joint modeling, inference methods, and issues. *Journal of Probability and Statistics*, 2012, 18. <https://doi.org/10.1155/2012/640153>
2. Ediebah, D. E., Galindo-Garre, F., Uitdehaag, B. M. J., Ringash, J., Reijneveld, J. C., Dirven, L., Zikos, E., Coens, C., van den Bent, M. J., Bottomley, A., & Taphoorn, M. J. B. (2015). Joint modeling of longitudinal health-related quality of life data and survival. *Quality of Life Research*, 24, 795–804. <https://doi.org/10.1007/s11136-014-0821-6>
3. Fairclough, D. L., Peterson, H. F., Cella, D., & Bonomi, P. (1998). Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. *Statistics in Medicine*, 17, 781–796. [https://doi.org/10.1002/\(SICI\)1097-0258\(19980315/15\)17:5/7%3c781::AID-SIM821%3e3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-0258(19980315/15)17:5/7%3c781::AID-SIM821%3e3.0.CO;2-O)
4. Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data*. Chapman and Hall/CRC.
5. Touraine, C., Cuer, B., Conroy, T., Beata, J., Gourgou, S., & Mollevi C. (2021). When a joint model should be preferred over a linear mixed model for analysis of longitudinal health-related quality of life data in cancer clinical trials. *BMC Medical Research Methodology* (under review).
6. Conroy, T., Galais, M.-P., Raoul, J.-L., Bouché, O., Gourgou-Bourgade, S., Douillard, J.-Y., Etienne, P.-L., Boige, V., Martel-Lafay, I., Michel, P., Llacer-Moscardo, C., François, E., Créhange, G., Abdelghani, M. B., Juzyna, B., Bedenne, L., & Adenis, A. (2014). Definitive chemoradiotherapy with FOLFOX versus fluorouracil and cisplatin in patients with oesophageal cancer (PRODIGE5/ACCORD17): Final results of a randomised, phase

- 2/3 trial. *The Lancet Oncology*, 15, 305–314. [https://doi.org/10.1016/S1470-2045\(14\)70028-2](https://doi.org/10.1016/S1470-2045(14)70028-2)
7. Bascoul-Mollevi, C. (2017). Health-related quality of life results from the PRODIGE 5/ACCORD 17 randomised trial of FOLFOX versus fluorouracil-cisplatin regimen in oesophageal cancer. *European Journal of Cancer*, 84, 239–249. <https://doi.org/10.1016/j.ejca.2017.07.038>
 8. Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., Haes, J. C. J. M. D., Kaasa, S., Klee, M., Osoba, D., Razavi, D., Rofe, P. B., Schraub, S., Sneeuw, K., Sullivan, M., & Takeda, F. (1993). The European organization for research and treatment of cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *JNCI Journal of the National Cancer Institute*, 85, 365–376. <https://doi.org/10.1093/jnci/85.5.365>
 9. Blazeby, J. M., Conroy, T., Hammerlid, E., Fayers, P., Sezer, O., Koller, M., Arraras, J., Bottomley, A., Vickery, C. W., Etienne, P. L., & Alderson, D. (2003). Clinical and psychometric validation of an EORTC questionnaire module, the EORTC QLQ-OES18, to assess quality of life in patients with oesophageal cancer. *European Journal of Cancer*, 39, 1384–1394. [https://doi.org/10.1016/S0959-8049\(03\)00270-3](https://doi.org/10.1016/S0959-8049(03)00270-3)
 10. James, G., Witten, D. M., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning*. Springer.
 11. Klein, J. P., & Moeschberger, M. L. (2005). *Survival analysis: Techniques for censored and truncated data, Second edition, corrected third printing*. Springer.
 12. Santos Nobre, J., & da Motta Singer, J. (2007). Residual analysis for linear mixed models. *Biometrical Journal*, 49, 863–875. <https://doi.org/10.1002/bimj.200610341>
 13. Verbeke, G., & Molenberghs, G. (2001). *Linear mixed models for longitudinal data Springer Series in Statistics*. Springer.
 14. Verbeke, G., Molenberghs, G., & Beunckens, C. (2008). Formal and informal model selection with incomplete data. *Statistical Science*, 23, 201–218. <https://doi.org/10.1214/07-STS253>
 15. Fitzmaurice, G., Laird, N., & Ware, J. (2011). *Applied longitudinal analysis*. Wiley.
 16. R Core Team. (2022). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>
 17. Rizopoulos, D. (2010). JM: An R package for the joint modeling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35, 1–33.
 18. Harrell, F. E. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis Springer Series in Statistics*. Springer International Publishing.
 19. Sauerbrei, W., Perperoglou, A., Schmid, M., Abrahamowicz, M., Becher, H., Binder, H., Dunkler, D., Harrell, F. E., Royston, P., Heinze, G., for TG2 of the STRATOS initiative. (2020). State of the art in selection of variables and functional forms in multivariable analysis—Outstanding issues. *Diagnostic and Prognostic Research*, 4, 3. <https://doi.org/10.1186/s41512-020-00074-3>
 20. Sauerbrei, W., Royston, P., & Binder, H. (2007). Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Statistics in Medicine*, 26, 5512–5528. <https://doi.org/10.1002/sim.3148>
 21. Royston, P., & Sauerbrei, W. (2008). *Multivariable model-building*. Wiley.
 22. Li, Z., Frost, H. R., Tosteson, T. D., Zhao, L., Liu, L., Lyons, K., Chen, H., Cole, B., Currow, D., & Bakitas, M. (2017). A semiparametric joint model for terminal trend of quality of life and survival in palliative care research. *Statistics in Medicine*, 36, 4692–4704. <https://doi.org/10.1002/sim.7445>
 23. Yang, L., Song, H., Peng, Y., & Tu, D. (2021). Joint analysis of longitudinal measurements and survival times with a cure fraction based on partly linear mixed and semiparametric cure models. *Pharmaceutical Statistics*, 20, 362–374. <https://doi.org/10.1002/pst.2082>
 24. Terrin, N., Rodday, A. M., & Parsons, S. K. (2015). Joint models for predicting transplant-related mortality from quality of life data. *Quality of Life Research*, 24, 31–39. <https://doi.org/10.1007/s11136-013-0550-2>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.