A Tutorial for Joint Modeling of Longitudinal and Time-to-Event Data in R

Sezen Cekic¹, Stephen Aichele¹,², Andreas M. Brandmaier³,⁴, Ylva Köhncke³,
Paolo Ghisletta¹,⁵,⁶

Abstract

In biostatistics and medical research, longitudinal data are often composed of repeated assessments of a variable and dichotomous indicators to mark an event of interest. Consequently, joint modeling of longitudinal and time-to-event data has generated much interest in these disciplines over the previous decade. In behavioural sciences, too, often we are interested in relating individual trajectories and discrete events. Yet, joint modeling is rarely applied in behavioural sciences more generally. This tutorial presents an overview and general framework for joint modeling of longitudinal and time-to-event data, and fully illustrates its application in the context of a behavioral study with the JMbayes R package. In particular, the tutorial discusses practical topics, such as model selection and comparison, choice of joint modeling parameterization and interpretation of model parameters. In the end, this tutorial aims at introducing didactically the theory related to joint modeling and to introduce novice analysts to the use of the JMbayes package.

Keywords
tutorial, joint model, mixed-effects model, time-to-event, association structures, JMbayes package, application
In many research settings, it is very common to record longitudinal observations that correspond to responses on continuous measures alongside dichotomous indicators to mark the occurrence of an event of interest. A prototypical example in clinical research is the repeated assessment of biological measures (e.g., blood pressure, antibody affinity, cholesterol level) that may relate to an event such as death, recovery from a disease, or disease diagnosis. In psychological research, much interest revolves around the association between long-term individual trajectories of cognitive performance and imminent death (e.g., terminal decline or terminal drop hypotheses; Kleemeier, 1962; Riegel & Riegel, 1972).

Methods for the separate analysis of such outcomes are well established in the literature, and these frequently include the use of mixed-effects models for the longitudinal portion of the data and the Cox proportional hazards (PH) model for the time-to-event data (Cox, 1972; Laird & Ware, 1982). However, the repeated measurements of the same individual are most likely related to the event, making them endogenous. This means that the measurement value at a given time point is informative about the future occurrence (or non-occurrence) of the event of interest. Likewise, the longitudinal trajectory is related to, and shaped by, the occurrence of the event. Separate analyses of such data cannot account for this endogeneity and provides no information about the associative strength between the two. A tentative solution to this problem is a two-step approach, whereby estimated parameters from the mixed-effects models are used as covariates in the time-to-event model. Yet, this approach has been shown to suffer from increased bias in estimation and loss of efficiency (Prentice, 1982; Rizopoulos, 2012). Joint models simultaneously estimate both longitudinal and time-to-event components and are better suited for analyzing such data because they estimate jointly the relative risk of the time-to-event outcome contingent upon the longitudinal outcome (Gould et al., 2015; Henderson, Diggle, & Dobson, 2000; Hogan & Laird, 1997; Papageorgiou, Mauff, Tomer, & Rizopoulos, 2019; Song, Davidian, & Tsiatis, 2002; Tsiatis & Davidian, 2004; Wulfsohn & Tsiatis, 1997).

Joint modeling is increasingly used in medical research (see for example Abdi et al., 2013; Brown & Ibrahim, 2003; He & Luo, 2016; Mauritz et al., 2011; Núñez et al., 2014; Rizopoulos & Takkenberg, 2014; Sène, Bellera, & Proust-Lima, 2014; Thabut et al., 2013). However, joint models remain conspicuously absent in psychological and behavioral research, despite the proliferation of data that could be analyzed more effectively using this powerful methodology, rather than by using a two-step estimation procedure. For example, given world-wide increases in human life expectancy (and corresponding increases in populations of elders), it will become increasingly important to understand how long-term changes in cognitive, personality, emotional, and other individual behavioral characteristics are related both to specific health-related outcomes (depression, stroke, dementia) and to longevity, or risk of death (World Health Organization, 2015). We found only a few behavioral studies using joint models to evaluate such associations (Ghisletta,
and we believe that the lack of an accessible and clear guidance for implementing such models remains an obstacle preventing non-expert users from applying them skillfully, or from adopting them at all.

This tutorial is therefore aimed at explaining basic features of the joint modeling framework for longitudinal and time-to-event data for users familiar with mixed-effects and time-to-event (a.k.a. survival) models. The overall aim is to guide readers step by step, such that at the end of the tutorial they acquire the basic concepts of joint modeling and are able to apply this methodology to their own research data. To do so, we initially reviewed the seven currently available joint-modeling packages in the open-source R environment for statistical computing (R Core Team, 2017). The seven packages are JM, JMbayes, joineR, lcmm, frailtypack, rstanarm and bamlss, which are described in Appendix in Tables A1, A2, A3, A4, A5, A6 and A7, respectively. We concluded that, at present, the JMbayes package (Rizopoulos, 2016) is the most comprehensive and extensible, and thus chose it as the basis for this tutorial. Note that a range of tools allowing one to perform joint modeling are available for other standard statistical software such as SAS/STAT, Stata, WinBUGS, and JAGS (see Papageorgiou et al., 2019, and references therein for more details).

JMbayes provides a framework wherein one or more characteristics from the longitudinal submodel can be flexibly parameterized as linear predictors in the time-to-event submodel. This flexibility is important because the relation between the continuous variable and the event of interest may take several functional forms, each of which may imply a different theoretical interpretation. For example, the longitudinal process can be specified in relatively simple terms (e.g., individual random effects about the intercept and the linear slope within a mixed-effects submodel), or in more complex non-linear spline effects, each of which can be linked to the time-to-event process with any of several association structures. JMbayes also allows for inclusion of multivariate longitudinal processes and for stratification criteria in the time-to-event submodel. Also, estimation of joint models, which are inherently complex, can be computationally intensive. Per its name, the JMbayes package relies on a Bayesian analytical approach, which is very attractive for applications where classical maximum likelihood estimation often encounters nonidentifiability issues yielding unreliable inferential conclusions. Indeed, Bayesian analysis using Markov chain Monte Carlo (MCMC) estimation can yield very precise point estimates without reliance on asymptotic assumptions (Robert & Casella, 2013). JMbayes also provides useful diagnostic measures for evaluating MCMC convergence and for evaluating relative model fit.

We note that the aim of this tutorial is not to replace the JMbayes documentation, which provides in-depth methodological detail and supplementary R code (Rizopoulos, 2016, 2020). Here, our goal is to provide a didactic introductory presentation to joint...
modeling, and to share computer syntax towards this aim, so that users may successfully carry out joint model analyses. For the sake of brevity and clarity, we do not provide an exhaustive treatment of all features of joint modeling from the applications, but we nevertheless briefly discuss some of them at the end of the tutorial, for those interested in advanced features. We begin by describing the statistical framework for joint modeling. We then discuss required software packages, describe the data set used for this tutorial, and finally present example analyses to illustrate the use of joint modeling for behavioral data analysis. We provide the data and the full syntax of our application in Supplementary Materials.

**Statistical Framework**

Below we describe the major features of the statistical theory behind joint modeling. For the sake of completeness, we show the most important equations of both the mixed-effects and the survival model, which lead to the joint model. Readers interested mainly in a conceptual understanding and in an empirical application of joint modeling may read the next paragraph, skip the remaining portions of this section (except the first paragraph of each subsection), and proceed to the next section.

The joint model, first described by Wulfsohn and Tsiatis (1997), consists of two submodels: a mixed-effects submodel for the longitudinal data and a time-to-event submodel for the time-to-event data. The two submodels are joined by allowing the time-to-event submodel to depend on some characteristics of the longitudinal submodel potentially in various ways. The defining feature of the joint model is that longitudinal and time-to-event data are modeled simultaneously with respect to a conditional joint density, rather than separately with two marginal and independent densities (see e.g., Rizopoulos, Hatfield, Carlin, & Takkenberg, 2014, for further technical details). Consequently, the effects of the longitudinal submodel on the time-to-event submodel are explicitly specified, and, at the same time, the parameter estimates of the former are also conditioned on the latter submodel (i.e., the longitudinal submodel is more accurately estimated by including the time-to-event data than without them, if endogeneity occurs). We discuss first the longitudinal submodel, then the time-to-event submodel, and finally several parametrization options to associate the two within the joint model.

**The Mixed-Effects Submodel**

Longitudinal data here consist of repeated measurements of the same outcome for each study participant over a given period of time. Repeated measurements for the same individual are typically correlated, and therefore each individual in the population is expected to have his or her own subject-specific response pattern over time. The mixed-effects model (Laird & Ware, 1982) accounts for this between person variability
by estimating person specific random effects around the parameters that are invariant across individuals (fixed effects).

Let \( y_i(t) \) denote the measurement on \( y \) for individual \( i \) (\( i = 1, \ldots, N \)) at time \( t \) (\( t = 1, \ldots, T_i^* \)), where \( T_i^* \) is the time-to-event for individual \( i \). The model assumes that \( y_i(t) \) is the observed measured value, which deviates from the true unobserved value \( \mu_i(t) \) by the amount of error \( \epsilon_i(t) \), where \( \epsilon_i(t) \sim \mathcal{N}(0, \sigma^2) \), normally distributed around 0 with variance \( \sigma^2 \). The longitudinal submodel can then be written as

\[
\mu_i(t) = X_i(t)\beta + Z_i(t)b_i, \quad \text{or} \quad \mu_i(t) = f(X_i(t)) + f_i(Z_i(t)),
\]

where \( b_i \) denotes the vector of random effects for individual \( i \) with \( b_i \sim \mathcal{N}(0, D) \), and \( D \) is the variance-covariance matrix of the random effects (typically a 2 x 2 unstructured matrix with random intercept and random slope effects). \( X_i(t) \) and \( Z_i(t) \) represent the design matrix containing the predictors for the fixed-effects regression coefficients \( \beta \) and for the random-effects regression coefficients \( b_i \), respectively (typically \( b_{i0} \) and \( b_{i1} \) for random intercept and slope effects).

The second line of Equation 1 represents the notation when \( \mu_i(t) \) is modeled non-parametrically via a spline approach (illustrated below). \( f(\cdot) \) and \( f_i(\cdot) \) represent the spline function for the fixed and, respectively, random components.

### The Time-to-Event Submodel

Time-to-event, or survival, analysis refers to statistical methods for analyzing time-to-event data. An event time is the time elapsed up to the occurrence of an event of interest (such as death or disease diagnostic), given that it has not previously occurred yet. Time-to-event data are characterized by the fact that the event of interest may not have occurred for every participant during the duration of the study. This particular type of missing data is analogously treated as data of a participant who drops out before the end of the study, and both scenarios constitute right censorship. Time-to-event analysis is particularly appropriate to analyze right-censored data.

There are different models for time-to-event data (Kaplan & Meier, 1958). Probably, the most popular is the semi-parametric Cox PH model (Cox, 1972), which is a regression model wherein covariates are assumed to have a multiplicative effect on the hazard for an event. The Cox PH model has the particularity of estimating the baseline hazard function non-parametrically. If there is a theoretical reason to suppose that the baseline hazard function has a particular parametric form (typically Weibull, Gamma or Exponential), a parametric model can be specified. Most often, however, an analyst ignores the exact form of a hazard function. The Cox PH model has been shown to approximate well unknown parametric hazard functions (Fox, 2002).

The Cox PH submodel can be formulated as
where \( h_0(t) \) is the baseline hazard function at time \( t \) and \( w_i \) is a vector of baseline predictors with corresponding regression coefficients \( \gamma \). The submodel has two distinct parts. First, \( h_0(t) \) describes the hazard (e.g., risk of the event occurring, given it has not yet occurred) when all covariates \( w_i \) have value zero. This baseline hazard function is assumed invariant across all individuals (\( h_0(t) \) does not depend on \( i \)). Second, the effect parameters \( \gamma^Tw_i \) describe how the hazard varies as a function of the explanatory covariates \( w_i \).

The Joint Model

Joint models formally associate the longitudinal and time-to-event processes through shared parameters (Henderson et al., 2000; Ibrahim, Chu, & Chen, 2010; Rizopoulos, 2012; Wulfsohn & Tsiatis, 1997). This framework therefore models the hazard of experiencing the focal event as dependent on a subject-specific characteristic of its longitudinal trajectory. More specifically, we have

\[
h_i(t) = h_0(t) \exp \left( \gamma^Tw_i \right),
\]

where \( h_0(t) \) is the baseline hazard function at time \( t \) and \( w_i \) is a vector of baseline predictors with corresponding regression coefficients \( \gamma \). The submodel has two distinct parts. First, \( h_0(t) \) describes the hazard (e.g., risk of the event occurring, given it has not yet occurred) when all covariates \( w_i \) have value zero. This baseline hazard function is assumed invariant across all individuals (\( h_0(t) \) does not depend on \( i \)). Second, the effect parameters \( \gamma^Tw_i \) describe how the hazard varies as a function of the explanatory covariates \( w_i \).

The “Current Value” Association

The first association structure between the longitudinal and the time-to-event submodels we discuss is known as the “current value”, and assumes that for individual \( i \), the true value \( \mu_i(t) \) of the longitudinal measure at time \( t \) is predictive of the risk \( h_i(t) \) of experiencing the event at that same time \( t \). The Cox PH submodel with this association structure is written as

\[
h_i(t) = h_0(t) \exp \left( \gamma^Tw_i \right),
\]

where \( h_0(t) \) is the baseline hazard function at time \( t \) and \( w_i \) is a vector of baseline predictors with corresponding regression coefficients \( \gamma \). The submodel has two distinct parts. First, \( h_0(t) \) describes the hazard (e.g., risk of the event occurring, given it has not yet occurred) when all covariates \( w_i \) have value zero. This baseline hazard function is assumed invariant across all individuals (\( h_0(t) \) does not depend on \( i \)). Second, the effect parameters \( \gamma^Tw_i \) describe how the hazard varies as a function of the explanatory covariates \( w_i \).
This equation associates $\mu_i(t)$, as defined in Equation 1, with $h_i(t)$ at each time point $t$. The reliable estimate of $\mu_i(t)$ is of chief importance, and is obtained by correctly specifying the design matrices $X_i(t)$ and $Z_i(t)$ in Equation 1. This means that we need to specify a longitudinal submodel that accurately describes the repeatedly observed measurements $y_i(t)$. The coefficient $\alpha_1$ provides a measure of the strength of the association between $\mu_i$ at time $t$ and the time-to-event process at the same time. More precisely, each one-unit increase on the current value of $\mu_i(t)$ is associated with an exp($\alpha_1$)-fold increase in a participant’s risk of experiencing the event at that same time $t$, given the event has not occurred prior to $t$. Note that this parameterization supposes that the associative parameter $\alpha_1$ between the longitudinal value and the event risk is the same across all individuals. However, the overall hazard $h_i(t)$ may vary across individuals as a function of the other subject-specific covariates $w_i$.

The “Current Value Plus Slope” Association

The “current value” parameterization, however, does not distinguish between individuals who have, at a specific time point, an equal longitudinal score (an equal $\mu_i(t)$), but who differ in their rate of change of this score (e.g., one subject may have an increasing trajectory, whereas another a decreasing trajectory). The second association structure between the time-to-event and the longitudinal submodels is known as the “current value plus slope” parameterization. This extends the previous structure by adding the rate of change of the measurement at time $t$, estimated as the derivative of $\mu_i(t)$ w.r.t. the time metric, as follows:

$$\mu_i'(t) = \frac{d\mu_i(t)}{dt},$$

where $\mu_i'(t)$ is the derivative of $\mu_i(t)$ with respect to the time variable $t$. Practically, the quantity $\mu_i'(t)$ has to be calculated “by hand” by the user. The reader is referred to any basic mathematical literature for calculus derivative rules. For worked examples, see Equation 13.

The corresponding submodel of this association structure has the form

$$h_i(t) = h_0(t) \exp \left( y^T w_i + \alpha_1 \mu_i(t) + \alpha_2 \mu'_i(t) \right),$$

where the hazard of experiencing the event at time $t$ is now assumed to be associated with both the value of $\mu_i(t)$ and its rate of change (i.e., slope) $\mu'_i(t)$, at time $t$ (see Equation 5). The coefficients $\alpha_1$ and $\alpha_2$ express the strength of the association between the value and rate of change of the (true) subject trajectory at time $t$ and the time-to-event process at the same time. For individuals having the same level of $\mu_i(t)$, the hazard ratio for
a one-unit increase of the current rate of change $\mu_i'(t)$ (i.e., the trajectory) is $\exp(\alpha_2)$. Identically, for individuals having the same rate of change $\mu_i'(t)$, the hazard ratio for a one-unit increase of the current value $\mu_i(t)$ is $\exp(\alpha_1)$.

**The “Shared Random-Effects” Association**

The third association structure between the time-to-event and the longitudinal submodels is a “shared random-effects” parameterization (Wulfsohn & Tsiatis, 1997). This parameterization includes only the random effects from the longitudinal submodel in Equation 1 as linear predictors of the relative risk submodel. For a simple random-effects structure (i.e., random intercept and slope effects), the random effects represent the individual deviations from the sample average intercept and slope values, and the association parameters therefore reflect the change in the log hazard for a one-unit change in these deviations. For more elaborate structures, such as nonlinear parametrization, associative parameters should be interpreted with caution (Gould et al., 2015; Rizopoulos, 2012; Rizopoulos et al., 2014). The “shared random-effects” parameterization postulates that, compared to the general tendency, individuals who have a lower (or higher) measure at baseline (i.e., random intercept) or who show a steeper decrease (or increase) in their longitudinal trajectories (i.e., random slope) are more or less likely to experience the event. The “shared random-effects” submodel can be written as

$$h_i(t) = h_0(t) \exp(\gamma^T w_i + \alpha^T b_i).$$

(7)

This parameterization is computationally simpler than the “current value” and the “current value plus slope” parameterizations, because the associative part $\alpha^T b_i$ in Equation 7 is time-independent. Indeed, when a simple linear random-intercept and/or random-slope structure is assumed for the longitudinal submodel in Equation 1, individual specificities are characterized by random effects $b_i$ which do not depend on time $t$. This means that for a specific individual $i$, random effects $b_i$ are unique, regardless of the time $t$, contrary to the current value $\mu_i(t)$ and the current slope $\mu_i'(t)$ (see e.g., Rizopoulos et al., 2014, for further details).

**Software**

The **JMbayes** package in R has a basic model fitting function called `jointModelBayes(·)`, which accepts as its main arguments a mixed-effects submodel object fitted by the function `lme(·)` from package `nlme` (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2017) and a time-to-event submodel object fitted by a Cox PH model with the function `coxph(·)` from the package `survival` (Therneau, 2020). Thus, both these packages need to be installed to use **JMbayes**. **JMbayes** also requires that Just Another Gibbs Sampler (JAGS; Plummer, 2004) be installed in order to perform MCMC estimation.
We show the R syntax to properly install the JMbayes package in Section Code 1 lines 1–4 of the Computer Code file in Supplementary Materials.

Both the nlme and survival packages are commonly used for fitting linear mixed-effects and, respectively, Cox PH submodels within the R environment. We hence do not discuss them at length here. We refer the interested readers to the original articles (Pinheiro et al., 2017; Therneau, 2020).

**Empirical Example**

We illustrate the joint modeling approach by applying JMbayes to data from a longitudinal study of cognitive change in adults (Rabbitt et al., 2004).

**The Manchester Longitudinal Study of Cognition Dataset**

The data come from the Manchester Longitudinal Study of Cognition (MLSC; Rabbitt et al., 2004). The MLSC study is a 20-plus-year study consisting in over 6200 adult individuals, with initial ages from 42 to 97. Five domains of cognitive ability were examined, but here, for simplicity, we focus only on processing speed (PS), defined as the speed at which we can compute simple but abstract tasks. This ability has been shown to decrease markedly during the lifespan, to induce general decline in more elaborate cognitive abilities (such as memory or verbal capacities), and to relate to physical health in older adults (Salthouse, 1996). Here, we investigate whether an individual’s level of and change in PS performance in adulthood predicts their risk of death.

PS was assessed up to four times at approximately 3-year intervals. Relations between cognition and survival have been the focus of several previous MLSC analyses (see Aichele, Rabbitt, & Ghisletta, 2015, and references therein for further details). For the current illustration, we use data from female participants aged between 50 and 90 years only. This gave an effective \( N = 2780 \) participants, of whom 1858 died during the study, who were tested on average twice. Other MLSC variables included in the current analyses were chronological age at each of the longitudinal assessments and at study entry for the longitudinal submodel, and smoking status and chronological age at study entry, time in the study, and survival status for the time-to event submodel.

Mixed-effects models require repeated data measured across time on each individual, and in our data set we have up to four individual measurements of the PS variable. There is no well-agreed-upon rule of thumb concerning the minimal amount of data required for fitting a longitudinal model properly, but the complexity of the fit (e.g., quadratic, cubic or non-parametric effects of the longitudinal variable) depends on the available amount of repeated data (e.g., Snijders, 2005; Snijders & Bosker, 2012). The quality of the longitudinal fit is directly related to the complexity of the model and its assessment is important. Indeed, the joint model directly includes parameter(s) of the longitudinal fit.
(current value, current slope, random effects) in the survival model. If the longitudinal model does not fit the repeated-measures data properly, the joint model will estimate invalid joint coefficients. We therefore advise the reader to check the longitudinal model quality (e.g., with the $R^2$ criterion) to ensure that the longitudinal model gives a reliable estimation of $\mu$ (see Equation 1), the longitudinal trajectory free of measurement error.

For the Cox model, the “1 in 10 rule” states that for every 10 occurrences of the event of interest, 1 predictor can be added to the model (Harrell, Lee, Califf, Pryor, & Rosati, 1984; Harrell, Lee, & Mark, 1996; Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996; Riley et al., 2019). In our data, we have 1858 deceased women (vs. 922 alive). In theory, then, according to the “1 in 10” rule, we could include as many as 185 predictors in our Cox model.

**MLSC Analysis Approach**

We hypothesize that mortality may explicitly depend on previous PS trajectories, whereby individuals with lower scores and/or steeper trajectories should have a greater hazard of dying than those performing better. We also posit that, given the PS-mortality association, adequate description of cognitive change in adult individuals should statistically take into account their (future) mortality information. A previous survival analysis of these data showed that the PH assumption was not met when comparing mortality risk for male and female participants (Aichele et al., 2015). Therefore, sex-specific analyses are warranted. For simplicity, we limit the current analyses to data from female MLSC participants. Before estimating the joint model, we perform initial model selection steps independently for the longitudinal and for the time-to-event submodels.

We first select a submodel for the longitudinal part of the data. Based on a previous literature search on the use of the mixed-effects model in cognitive aging research (Ghisletta et al., in press), we tested six submodels: the first with fixed and random-effects for the intercept and the linear slope (to describe change over time), the second adding to the first a quadratic slope (change trajectories with a single curvature), the third adding to the second a cubic slope (change trajectories with two curvatures), and the fourth to the sixth with spline effects of time and with different numbers of nodes. Of the six submodels, based on statistical criteria, we choose the one that provides the most satisfying description of the PS trajectories.

We then, separately, select the optimal time-to-event submodel. To do so we rely on the Cox PH model that relates smoking status and initial age to the timing and occurrence of death. For each covariate we test the PH assumption and then retain the submodel that best describes the mortality process based on the selected covariates. Finally, we combined both submodels in joint modeling, and compared four association structures between the two submodels.

The choice of including or not given predictors in the mixed-effects and in the Cox submodels should be theoretically driven. For the longitudinal model, the objective is
to describe the longitudinal trajectory across time of the variable of interest for each individual accounting for measurement noise. Besides the obvious choice of including time, at both the overall and subject-specific level, any other predictor that might improve the prediction accuracy of the outcome should be added. Likewise, for the time-to-event submodel, any predictor that might help forecasting the hazard of death should be considered for inclusion. Again, time is a natural predictor here, too. Finally, if theoretically meaningful, given predictors can be included in both the mixed-effects and in the time-to-event submodels, so that they might ameliorate both outcomes.

Longitudinal Model Selection

As required by the JMbayes package, we fit the longitudinal submodel with the \texttt{lme(\cdot)} function from the \texttt{nlme} package (Pinheiro et al., 2017). This longitudinal submodel must be specified in terms of its fixed and random components. For both components, linear and/or higher-degree polynomial terms are allowed (e.g., linear, quadratic or cubic change). More complex parameterizations of nonlinear effects are also feasible, such as achieved through the flexible spline methodology (see the second line in Equation 1 and cf. Rizopoulos, 2012).

The spline approach has gain much popularity in recent years because of its flexibility in describing subject-specific longitudinal trajectories that follow highly nonlinear functions and that, individually, deviate from the average sample function. A spline is a piecewise function composed of polynomials adjusted to data within adjacent intervals, separated by equidistant points called nodes. The degree of the spline is defined by the polynomial of the highest degree used. For instance, if the function fitted to the data within an interval is a straight line, the spline is said to be of degree 1. Cubic splines are a popular specification for splines that entails polynomials of degree 3 (Hastie & Tibshirani, 1987). Within the \texttt{lme(\cdot)} function, nonlinear effects can be specified with the subfunction \texttt{ns(\cdot)} from the splines package. The \texttt{ns(\cdot)} function uses natural cubic splines that estimate a smooth functions of the predictors (functions $f(\cdot)$ and $f_1(\cdot)$ in the second line of Equation 1; Chambers, 1991). With natural cubic splines, the data within each interval are fitted by a cubic spline, and splines from neighboring intervals must respect two conditions: at each node, the first and the second derivatives of the splines are continuous, meaning that, for any two adjacent intervals, the finishing portions of the left spline smoothly joins the starting portion of the right spline, thereby giving the visual impression of a single, continuous line passing through both intervals.

The user must select the optimal number of nodes, which is a rather arbitrary component of splines. However, this value should be less than or equal to the number of repeated measures. To this end, there is no automatic selection procedure implemented in the JMbayes package (in contrast to the mgcv package, for example; Wood, 2006). We suggest to first investigate the number of nodes for the fixed part of the longitudinal submodel, before proceeding to explore that of random part (when both components are
modeled using splines, as proposed in Rizopoulos, 2016). To do so we suggest testing multiple submodels with different number of nodes and comparing them on the basis of their Bayesian Information Criterion (BIC), where smaller values are preferable (Schwarz, 1978). The BIC is available in the \texttt{lme(·)} output and is known to penalize complex models more strongly than the Akaike Information Criterion (Akaike, 1974). Conveniently, the BIC allows a fully Bayesian interpretation (Raftery, 1995). The difference in BIC between two models is to be interpreted as “not worth mentioning,” “positive,” “strong,” and “very strong” evidence against the model with highest value when the difference is between 0 and 2, 2 and 6, 6 and 10, and greater than 10, respectively. (Kass & Raftery, 1995, p. 777). Applications of the spline-based approach within the joint modeling framework can be found, for example, in Rizopoulos, Verbeke, and Lesaffre (2009), Brown, Ibrahim, and DeGruttola (2005), Ding and Wang (2008), and Rizopoulos and Ghosh (2011).

As explained in Oehlert (2012) and Pinheiro and Bates (2000), likelihood ratio tests to compare models with different fixed effects and same random effects should be performed using maximum likelihood estimation (ML). However, the default estimation method implemented within the \texttt{lme(·)} function uses Restricted Maximum Likelihood (REML), because REML, unlike ML, produces unbiased estimates of variance and covariance parameters (the difference in bias vanishes with high number of observations). To meaningfully compare longitudinal submodels in terms of their BIC they should thus be re-estimated with ML.

For this illustration, we modeled change in cognitive performance as a function of age in years centered at age 50 (\texttt{Age}), conditioned on age in years at study entry (\texttt{AgeStart}). Thus, we add \texttt{AgeStart} and \texttt{Age} as fixed parameters to the longitudinal submodel. We also modeled \texttt{Age} as a random effect, because visually we might infer differences in rates of decline. Individual observed trajectories of PS are depicted in Figure 1.

We estimated the following specifications of the longitudinal submodel:

\begin{align*}
\text{lmeFit.mlsc1: } y_i(t) &= (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})\text{Age}_i(t) + \beta_4\text{AgeStart}_i + \epsilon_i(t), \\
\text{lmeFit.mlsc2: } y_i(t) &= (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})\text{Age}_i(t) + \beta_2\text{Age}_i(t)^2 + \beta_4\text{AgeStart}_i + \epsilon_i(t), \\
\text{lmeFit.mlsc3: } y_i(t) &= (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})\text{Age}_i(t) + \beta_2\text{Age}_i(t)^2 + \beta_3\text{Age}_i(t)^3 + \beta_4\text{AgeStart}_i + \epsilon_i(t), \\
\text{lmeFit.mlsc4: } y_i(t) &= f(\text{Age}_i(t), 2) + b_{0i} + b_{1i}\text{Age}_i(t) + \epsilon_i(t), \\
\text{lmeFit.mlsc5: } y_i(t) &= f(\text{Age}_i(t), 3) + b_{0i} + b_{1i}\text{Age}_i(t) + \epsilon_i(t), \\
\text{lmeFit.mlsc6: } y_i(t) &= f(\text{Age}_i(t), 4) + b_{0i} + b_{1i}\text{Age}_i(t) + \epsilon_i(t).
\end{align*}
The first three \( \text{lmeFit.mlsc1}, \text{lmeFit.mlsc2}, \text{lmeFit.mlsc3} \) estimate, respectively, a polynomial of first, second, and third degree of Age, respectively, with random effects for the intercept and Age. Models 4, 5, 6 \( \text{lmeFit.mlsc4}, \text{lmeFit.mlsc5}, \text{lmeFit.mlsc6} \) estimate cubic spline effects of Age with, respectively, 2, 3 and 4 nodes. Here, too, we included random effects for the intercept and Age. To do so, we used the following R code:

```r
ctrl <- lmeControl(opt="optim")

lmeFit.mlsc1 <- lme(FS_SPD ~ Age_50+AgeStart, 
data = MLSC_fem2, random = ~ Age_50 | ID)
lmeFit.mlsc2 <- lme(FS_SPD ~ Age_50+I(Age_50^2)+AgeStart, 
data = MLSC_fem2, random = ~ Age_50 | ID)
lmeFit.mlsc3 <- lme(FS_SPD ~ Age_50+I(Age_50^2)+I(Age_50^3)+AgeStart, 
data = MLSC_fem2, random = ~ Age_50 | ID)
lmeFit.mlsc4 <- lme(FS_SPD ~ ns(Age_50,2)+AgeStart, data = MLSC_fem2, 
random = ~ Age_50 | ID)
lmeFit.mlsc5 <- lme(FS_SPD ~ ns(Age_50,3)+AgeStart, data = MLSC_fem2, 
random = ~ Age_50 | ID)
lmeFit.mlsc6 <- lme(FS_SPD ~ ns(Age_50,4)+AgeStart, data = MLSC_fem2, 
random = ~ Age_50 | ID)

m1<-update(lmeFit.mlsc1, method = "ML")
```

Figure 1

Subject-Specific Longitudinal Trajectories of PS Cognitive Performance for 100 Randomly Selected Women in the MLSC Database With and Without an Event
m2<-update(lmeFit.mlsc2, method = "ML")
m3<-update(lmeFit.mlsc3, method = "ML")
m4<-update(lmeFit.mlsc4, method = "ML")
m5<-update(lmeFit.mlsc5, method = "ML")
m6<-update(lmeFit.mlsc6, method = "ML")

lmeFit.mlsc2.1 <- lme(FS_SPD ~ Age_50*AgeStart+I(Age_50^2),
data = MLSC_fem2, random = ~ Age_50 | ID)
lmeFit.mlsc2.2 <- lme(FS_SPD ~ Age_50*AgeStart+I(Age_50^2)*AgeStart,
data = MLSC_fem2, random = ~ Age_50 | ID)

Note that ctrl <- lmeControl(opt="optim") changes the optimizer to be used
for lme function (nlminb is the default).

We evaluated the adjustments of the six submodels in terms of their respective BIC
values. The R code to this end is

BIC(m1,m2,m3,m4,m5,m6)

and the results are displayed in Table 1.

Table 1

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>7</td>
<td>29414.87</td>
</tr>
<tr>
<td>m2</td>
<td>8</td>
<td>29158.73</td>
</tr>
<tr>
<td>m3</td>
<td>9</td>
<td>29167.24</td>
</tr>
<tr>
<td>m4</td>
<td>8</td>
<td>29162.86</td>
</tr>
<tr>
<td>m5</td>
<td>9</td>
<td>29170.97</td>
</tr>
<tr>
<td>m6</td>
<td>10</td>
<td>29178.68</td>
</tr>
<tr>
<td>m2.1</td>
<td>9</td>
<td>29155.24</td>
</tr>
<tr>
<td>m2.2</td>
<td>10</td>
<td>29163.72</td>
</tr>
</tbody>
</table>

Note. BIC = Bayesian Information Criterion.

For this we concluded that the submodel with a degree-2 polynomial for the fixed effects
of Age and with a random intercept and a random Age slope fit the data best (m2,
corresponding to lmeFit.mlsc2).

We subsequently expanded this submodel by adding interaction effects between the
degree-2 Age polynomial and AgeStart, as shown in Equation 9. We again evaluated
statistical fit with the BIC. The corresponding R code and related equations are

lmeFit.mlsc2.1 <- lme(FS_SPD ~ Age_50*AgeStart+I(Age_50^2),
data = MLSC_fem2, random = ~ Age_50 | ID)
lmeFit.mlsc2.2 <- lme(FS_SPD ~ Age_50*AgeStart+I(Age_50^2)*AgeStart, 
data = MLSC_fem2, random = ~ Age_50 | ID)
m2.1<-update(lmeFit.mlsc2.1, method = "ML" )
m2.2<-update(lmeFit.mlsc2.2, method = "ML" )
BIC(m2,m2.1,m2.2)

The comparison between these two submodels and the previous m2 submodel is also displayed in Table 1 and indicates that the submodel m2.1 (corresponding to lmeFit.mlsc2.1) best fit the data (the R code can be found in Section Code 3 of the Computer Code file in Supplementary Materials).

To ensure that our model fits the data properly, we need to calculate the $R^2$ criterion relative to our selected longitudinal model. Note that the $R^2$ criterion for linear mixed-effects models can be obtain following Nakagawa and Schielzeth (2013) with the following R commands:

library(MuMIn)
r.squaredGLMM(lmeFit.mlsc2.1)

The first column of the output contains the marginal $R^2$, the proportion of variance explained by the fixed effects only, whereas the second column shows conditional $R^2$, the proportion of variance explained by both the fixed and random effects (the quantity of interest for us). In our case, the conditional $R^2$ equals .983, which means that the proportion of variance explained in our final longitudinal model is 98.3%. We therefore have a model with an excellent fit to pursue our joint model estimation procedure.

To conclude this section on the longitudinal submodel of the MLSC PS score, we retain submodel lmeFit.mlsc2.1, which will be joined to the upcoming time-to-event submodel in the final joint model. The function for the true value of PS $\mu_i(t)$ (where the observed PS score $y_i(t) = \mu_i(t) + \epsilon_i(t)$) is

$$
\mu_i(t) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})Age_i(t) + \beta_2Age_i(t)^2 + \beta_3AgeStart_i + \beta_4AgeStart_i \times Age_i(t).
$$

Cekic, Aichele, Brandmaier et al. 15
Quantitative and Computational Methods in Behavioral Sciences
2021, Article e2979, https://doi.org/10.5964/qcmb.2979
Time-to-Event Model Selection

As discussed previously, the Cox PH submodel, defined in Equation 2, is often the preferred choice to analyze time-to-event data, given it does not require the user to specify a parametric baseline hazard function $\lambda_0(t)$. The nonparametric specification of the baseline hazard function within the `coxph(·)` function in R uses a B-spline approach, but if desired, regression splines can instead be invoked by setting the `base-Haz` function argument accordingly. The number and the position of the nodes are automatically selected, but they can also be fixed via the `lns.in.kn` and the `knots` function arguments.

In this illustration we specified a nonparametric baseline hazard function within the Cox PH submodel, with factor Smoker (whether or not an individual was a smoker at study entry) as baseline variable, conditioned on age in years at study entry (AgeStart). The corresponding R code is

```r
Coxfit_fem <- coxph(Surv(AgeLastObserved_2012_50, Dead_By2012) ~ Smoker+AgeStart, data = MLSC_ID_fem2, x=TRUE)
```

Note that in the joint modeling context we need to set `x = TRUE` (or equivalently `model = TRUE`) in the call of the `coxph(·)` function so that the design matrix used in the Cox model is returned in the object fit (required for the joint model). Importantly, note that the scale of the temporal variable (here Age and AgeLastObserved_2012_50) should be identical in the longitudinal and in the survival models. The indicator for the event (alive or dead) is Dead_By2012.

Results are displayed in Table 2. The results show that factor Smoker is significant ($p < .001$). The estimated coefficient means that being smoker at study entry is associated with a hazard ratio of $\exp(0.519) = 1.680$. That is, being smoker at study entry increases the relative risk of death by 68%, compared to a non-smoker at entry.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>coef</th>
<th>exp(coef)</th>
<th>se(coef)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoker</td>
<td>0.519</td>
<td>1.680</td>
<td>0.064</td>
<td>8.092</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>AgeStart</td>
<td>0.099</td>
<td>1.104</td>
<td>0.004</td>
<td>9.815</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

An important (and often overlooked) assumption of the Cox PH model is that baseline hazard functions for model predictors are proportional (Cox, 1972). In other words, survival curves for different strata of a given predictor (determined by the particular choices of values for the $\gamma$-variables), have hazard functions that are proportional over
time. For example, if the model specifies sex (male vs. female) as a mortality predictor, it is assumed that the hazard functions for males and females are proportional.

The PH assumption can be checked statistically using the Schoenfeld Residuals Test (SRT; Schoenfeld, 1982), which evaluates independence between model residuals and the time variable (here Age) separately for each covariate. These residuals represent the difference between the observed covariate (for each subject and each covariate separately) and the expected value of the covariate given the risk set at that time (see Schoenfeld, 1982, for further technical details). More precisely, the SRT checks whether the correlation between the scaled residuals and the time variable differs significantly from zero for each predictor included in the model. If so, the residuals are thought to be related to time, thereby violating the PH assumption, because the hazard changes in time.

The SRT can be applied for each predictor individually and also for the entire model with all predictors, using the function `cox.zph(·)` from the `survival` package. Table 3 displays the results of the SRT, and the following R code allows to perform the SRT test:

```r
print(cox.zph(Coxfit_fem))
```

<table>
<thead>
<tr>
<th>Predictor</th>
<th>rho</th>
<th>chisq</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoker</td>
<td>0.068</td>
<td>8.38</td>
<td>.004</td>
</tr>
<tr>
<td>AgeStart</td>
<td>0.061</td>
<td>6.34</td>
<td>.012</td>
</tr>
<tr>
<td>GLOBAL</td>
<td>N/A</td>
<td>15.81</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

The results of the SRT indicate that the PH assumption does not hold for the factor Smoker and for AgeStart as a predictor ($p < .05$), nor for the whole submodel ($p < .05$). When the PH assumption is not met for a predictor of theoretical interest, a stratification procedure for that predictor is required (Fox, 2002). This stratification procedure is supported within the `coxph(·)` function inside the `JMbayes` package. In our particular case, we should thus stratify the time-to-event analysis by smoker status, with two strata, smoker vs. non smoker, and also initial age, with an arbitrary number of strata (e.g., 50–59, 60–69, ..., 91–99 years), where within each the PH assumption is to be met. For the sake of simplicity within this tutorial, we do not use the stratification procedure and maintain the variables Smoker and AgeStart in the Cox PH submodel, as though the PH assumption were met for both. Kaplan-Meier survival curves with respect to Age by smoker status are depicted in Figure 2, and the summary of those survival curves can be obtained with the following commands:
fit<-survfit(Surv(AgeLastObserved_2012_50, Dead_By2012) ~ Smoker, MLSC_ID_fem2, id=ID)
print(fit)

**Table 4** displays the output of the summary:

<table>
<thead>
<tr>
<th>Factor</th>
<th>n</th>
<th>Events</th>
<th>Median centered age at death</th>
<th>0.95 LCL</th>
<th>0.95 UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoker=FALSE</td>
<td>2384</td>
<td>1562</td>
<td>39</td>
<td>38</td>
<td>39</td>
</tr>
<tr>
<td>Smoker=TRUE</td>
<td>396</td>
<td>296</td>
<td>34</td>
<td>33</td>
<td>35</td>
</tr>
</tbody>
</table>

*Note.* LCL = Lower confidence limit; UCL = Upper confidence limit.

**Figure 2**

*Kaplan-Meier Estimator of Survival Probabilities for Smokers vs. Non-Smokers*

*Note.* Age is centered at 50 years.

**Joint Model**

We are now ready to explore the association between longitudinal trajectories in processing speed and risk of death. We consider the three association structures as discussed above and, again, use a model comparison approach to select the final model.

The **“current value” association** — We start with the “current value” parameterization defined in Equation 4. For our data, this joint model is written as
\[ h_i(t) = h_0(t) \exp \left( \beta_1 \text{Smoker}_i + \beta_2 \text{AgeStart}_i + \alpha_i \mu_i(t) \right), \quad (11) \]

with \( h_0(t) \) denoting the baseline hazard function. We estimated this submodel with the R command

```r
jointFit.mlsc1 <- jointModelBayes(lmeFit.mlsc2.1, Coxfit_fem, 
timeVar = "Age_50", n.iter = 30000)
```

with \( \text{lmeFit.mlsc2.1} \) the longitudinal submodel, \( \text{Coxfit_fem} \) the survival Cox submodel and \( \text{timeVar = "Age_50"} \) the temporal variable \( \text{Age_50} \). Note that no specification of the association structure is necessary, because the “current value” parameterization is the default implementation. Here we specified that, after the burn-in period, the MCMC algorithm should run for 30000 iterations.

Additionally, in a Bayesian paradigm, prior distributions must be specified for model parameters. A basic approach is to simply select non-informative priors, which is the default for \text{JMbayes}. If desired, informative prior distributions can also be specified. For example, in Andrinopoulou and Rizopoulos (2016), where the goal was to estimate elaborate models with many predictors in the time-to-event submodel, shrinkage priors were used in order to force small effects to be equal to zero while maintaining true large effects (van Erp, Oberski, & Mulder, 2019). Of importance, \text{JMbayes} uses parameter estimates from the individual longitudinal and time-to-event submodels, as illustrated above, as starting values for the MCMC estimation of the joint model. Also, the package relies on a single, long chain, rather than multiple parallel, short chains.

As usual, before interpreting the results, we should worry about the estimation quality of the model. Various diagnostic plots are available within \text{JMbayes}. The \text{plot(·)} function can be used to visually examine convergence of the \text{jointModelBayes(·)} MCMC estimation. The output includes trace plots, autocorrelation plots, and kernel density estimation plots. An example of these plots for the association parameter fitted within the joint model with current value parametrization can be found in Figure 3.

MCMC samples are dependent by construction. This particularity does not impact the validity of the posterior distribution estimation if the algorithm has enough iterations to explore this posterior distribution. To obtain the same Monte Carlo error for an estimation, correlated MCMC estimation needs more samples. In the trace plots, we seek the so-called “lazy caterpillar” pattern, which means that across the various iterations the estimated values of the parameters are within reasonably restricted ranges, and are not suddenly outside such a range. For the autocorrelation plots, which represent the correlations between a solution and that a given lag from it, we seek autocorrelations that become small and as close to zero as quickly as possible (i.e., with small lags), meaning that the solutions of the simulated samples become quickly independent. The kernel density plots are smoothed histograms of the estimated solutions across the simulated samples. We expect these to be unimodal and with small tails. It is also possible
to specify MCMC estimation with multiple shorter chains, which can also be diagnosed, but this requires additional R coding: Users must select different initial values within the priors space and then extract the corresponding MCMC output for analysis using the coda MCMC diagnostic package (Plummer, Best, Cowles, & Vines, 2006; Rizopoulos, 2016).

In our case, the diagnostic plot in Figure 3 shows a positive autocorrelation in the trace plot (top of the Figure 3) and an autocorrelation that remains significant around .2 at lag 30 (second line of Figure 3). We therefore asked for 30000 iterations (enough to obtain a valid posterior distribution estimation) with a burn-in period of 3000 (the algorithm adapts during the first 3000 iterations). JMbayes also implements an adaptive
thinning (the lag between two iterations taken into account in the estimation) so that the maximum number of samples in memory does not exceed 2000. Here, the thinning is therefore 15 (30000/15 = 2000).

The estimated parameter values can be obtained with the standard `summary(·)` function. The R output of `summary(jointFit.mlsc1)` can be found in the Section Code 2 of the Computer Code file in Supplementary Materials. We show the results of this first association structure in Table 5. We report the posterior means of all estimated parameters, with their 95% credible intervals. The Assoct coefficient in the output represents \( \alpha \) in Equation 11 and estimates the association between the true processing speed score at time \( t \) and the risk of death at that same time. We can interpret this parameter as follows: The percent reductions in mortality risk associated with a one-unit increase in the current true value of processing speed for women is calculated as \((1 - HR) \times 100\), where HR is the hazard ratio corresponding to a one-unit difference in the current true value of processing speed \( \mu_i(t) \). Thus, HR is calculated as \( \exp(-0.026) = 0.974 \), which gives \( 1 - 0.974 = 0.026 \), and translates to 2.6% lower risk of death per one unit increase in the current true value of processing speed. The effect size is small but the parameter is highly significant: it means that a decrease of one point in true speed score is not very important, even though a lower true PS score has a strong impact on survival.

### Table 5

*Parameter Estimates and 95% Credible Intervals for the “Current Value” Association*

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Value</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoker</td>
<td>0.445</td>
<td>0.328</td>
<td>0.555</td>
</tr>
<tr>
<td>AgeStart</td>
<td>-0.048</td>
<td>-0.055</td>
<td>-0.038</td>
</tr>
<tr>
<td>Assoct</td>
<td>-0.026</td>
<td>-0.032</td>
<td>-0.020</td>
</tr>
<tr>
<td>tauBs</td>
<td>12.186</td>
<td>4.643</td>
<td>25.541</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>8.806</td>
<td>6.506</td>
<td>11.060</td>
</tr>
<tr>
<td>Age</td>
<td>0.192</td>
<td>0.079</td>
<td>0.302</td>
</tr>
<tr>
<td>AgeStart</td>
<td>-0.054</td>
<td>-0.090</td>
<td>-0.018</td>
</tr>
<tr>
<td>I(Age^2)</td>
<td>-0.007</td>
<td>-0.007</td>
<td>-0.006</td>
</tr>
<tr>
<td>Age:AgeStart</td>
<td>-0.005</td>
<td>-0.006</td>
<td>-0.003</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.946</td>
<td>0.916</td>
<td>0.977</td>
</tr>
<tr>
<td>D[1, 1]</td>
<td>31.550</td>
<td>28.805</td>
<td>34.265</td>
</tr>
<tr>
<td>D[2, 1]</td>
<td>-0.320</td>
<td>-0.402</td>
<td>-0.229</td>
</tr>
<tr>
<td>D[2, 2]</td>
<td>0.078</td>
<td>0.073</td>
<td>0.084</td>
</tr>
</tbody>
</table>

*Note.* D[i, j] = ij-element of the covariance matrix of the random effects; tauBs = hyperparameter relative to the distribution of the baseline hazard function \( h_0(t) \) and is typically not interpreted.
The coefficient for smoker status (0.445) indicates a strong association with mortality, with an exp(0.445) = 1.560-fold increase in risk of death in smokers, compared to non-smokers. The intercept value tells us that when the covariates Age and AgeStart are equal to zero (equivalent to 50 years for subjects’ age), the average PS value is 8.806. In addition D[1,1] shows the variance of the intercept across subjects, D[2,2] is the variability in the slope across subjects, and D[2,1] indicates the negative correlation between random intercept and random slope effects.

The “current value plus slope” association — This second association structure takes the form

\[ h_i(t) = h_0(t) \exp \left( \beta_1 \text{Smoker}_i + \beta_2 \text{AgeStart}_i + \alpha_1 \mu_i(t) + \alpha_2 \mu'_i(t) \right). \quad (12) \]

and specifies that the hazard of experiencing the event at time \( t \) is now assumed to be associated both with the true value of processing speed at time \( t, \mu_i(t) \), and its current rate of change (i.e., slope) at time \( t \), noted \( \mu'_i(t) \).

Recalling Equation 10, the temporal variable is Age. The derivative of Equation 10 with respect to the temporal variable Age can be written as

\[ \mu'_i(t) = \frac{\partial \mu_i(t)}{\partial \text{Age}} = \beta_1 + b_1 + \beta_2 (2 \times \text{Age}_i(t)) + \beta_4 \text{AgeStart}_i. \quad (13) \]

The R code for specifying the derivative form of this linear submodel and for running the corresponding joint model is

```r
dForm <- list(fixed= ~1 + I(2*Age_50) + AgeStart, random = ~ 1, indFixed = c(2,4,5), indRandom = 2)
```

This R code means that the derivative of the current value (\( \mu_i(t) \)), with respect to Age_50, is equal for the fixed part (fixed=) to an intercept plus 2 times the Age_50 variable plus AgeStart (\( \beta_1 + \beta_2 (2 \times \text{Age}_i(t)) + \beta_4 \text{AgeStart}_i \)) in Equation 13. The derivative of the random part of the model in Equation 13, with respect to Age_50, equals an intercept (\( b_1 \) in Equation 13).

Arguments indFixed=c(2,4,5) and indRandom=2 refer to indices of the fixed and random model parameters corresponding to the retained (non-zero) arguments in the derivative of Equation 13. Thus, for the fixed parameters, indices c(2,4,5) relate to row numbers in the output of `summary(jointFit.mslc1)` under Longitudinal Process that correspond to the appropriate fixed model parameters \( \{\beta_1, \beta_2, \beta_4\} \) (see Computer Code Section Code 4, lines 31–37). Similarly, for the random parameters,
index 2 relates to row number relative to the random parameter $b_1$, found under \textit{Variance Components} (see Computer Code Section Code 4, lines 24–28).

We show the results of the “current value plus slope” parametrization in Table 6 and the R code allowing us to estimate the model is

```r
jointFit.mlsc2 <- update(jointFit.mlsc1, param = "td-both", extraForm = dForm)
```

Table 6
\textit{Parameter Estimates and 95\% Credible Intervals for the “Current Value Plus Slope” Association}

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Value</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoker</td>
<td>0.424</td>
<td>0.315</td>
<td>0.573</td>
</tr>
<tr>
<td>AgeStart</td>
<td>-0.056</td>
<td>-0.066</td>
<td>-0.048</td>
</tr>
<tr>
<td>Assoct</td>
<td>0.001</td>
<td>-0.008</td>
<td>0.010</td>
</tr>
<tr>
<td>AssoctE</td>
<td>-1.600</td>
<td>-1.984</td>
<td>-1.243</td>
</tr>
<tr>
<td>tauBs</td>
<td>14.250</td>
<td>5.747</td>
<td>28.615</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>8.502</td>
<td>6.232</td>
<td>10.765</td>
</tr>
<tr>
<td>Age</td>
<td>0.185</td>
<td>0.070</td>
<td>0.295</td>
</tr>
<tr>
<td>AgeStart</td>
<td>-0.049</td>
<td>-0.086</td>
<td>-0.012</td>
</tr>
<tr>
<td>I(Age$^2$)</td>
<td>-0.007</td>
<td>-0.007</td>
<td>-0.006</td>
</tr>
<tr>
<td>Age:AgeStart</td>
<td>-0.005</td>
<td>-0.006</td>
<td>-0.003</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.946</td>
<td>0.910</td>
<td>0.981</td>
</tr>
<tr>
<td>D[1, 1]</td>
<td>31.093</td>
<td>28.992</td>
<td>33.549</td>
</tr>
<tr>
<td>D[2, 1]</td>
<td>-0.301</td>
<td>-0.396</td>
<td>-0.216</td>
</tr>
<tr>
<td>D[2, 2]</td>
<td>0.079</td>
<td>0.073</td>
<td>0.085</td>
</tr>
</tbody>
</table>

\textit{Note.} D[i, j] = ij-element of the covariance matrix of the random effects; \textit{tauBs} = hyperparameter relative to the distribution of the baseline hazard function $h_0(t)$ and is typically not interpreted.

Assoct represents $\alpha_1$ and AssoctE represents $\alpha_2$ in Equation 6, and quantify, respectively, the association between the current true value and the current rate of change of processing speed and the relative risk of death. We obtain $\alpha_1 = 0.001$ and $\alpha_2 = -1.600$, with the 95\% credible interval of Assoct that contains 0, while that of AssoctE does not. Clearly, then, the rate of change in perceptual speed at a given time point is a much more important predictor of than the actual true value of speed at that same time point. The lower the rate of change, the greater the odds of dying.

To interpret the effect of AssoctE on the risk of death, we could proceed as usual, by relating it to a one-unit change. However, whereas a one-unit value is a meaningful comparison for the current true value of speed, it is excessively large for the rate of change. To show this, we plotted in Figure 4 the current true value $\mu(t)$ estimated from the previous joint model with the “current value” association (continuous line) and the
current rate of change \( \mu(t) \) (dashed line) from the joint model “current value plus slope”, as a function of the Age variable. We can immediately see that whereas the current value changes over a large range (from almost 6 to \(-10\)), the rate of change varies much less (from about 0 to less than \(-1\)). Hence, whereas a one-unit change could be useful in interpreting the former effect, it is less so for the latter effect (it would result in a \((1 - \exp(-1.600)) \times 79.81\%\) reduction in mortality risk, which is simply huge). Coming back to Figure 4, we note that over an interval of 40 years, the current true value decreases by 14.722 points, whereas the rate of change decreases by only 0.535 points. This translates to an average decrease over a 10-year period of 3.684 for the current true value and 0.134 for the rate of change, with associated percent reductions in mortality risk of 9.13\% (calculated as \((1 - \exp(-0.026 \times 3.684)) \times 100\)) and 19.3\% (calculated as \((1 - \exp(-1.6 \times 0.133)) \times 100\)), respectively. We believe this exercise to be most useful to properly interpret the effect sizes stemming from this model, because, typically, the range of the current true value is much larger than that of its rate of change.
The “shared random effects” association — The third joint model can be written as

\[ h_i(t) = h_0(t) \exp \left( \beta_1 \text{Smoker}_i + \beta_2 \text{AgeStart}_i + \alpha_3 b_{0i} + \alpha_4 b_{1i} \right), \]  

(14)

where \( \alpha_3 \) and \( \alpha_4 \) are the parameters of association of \( b_{0i} \) and \( b_{1i} \), which are defined in Equation 8 as the individual deviations from the sample average intercept and slope values, respectively. \( \alpha_3 \) and \( \alpha_4 \) define the association between the random intercept and the random slope effects, respectively, and the hazard of death. More precisely, for individuals having the same deviation from the average slope, the hazard ratio for a one-unit increase in individual deviation from the average intercept is \( \exp (\alpha_3) \). Conversely, for individuals having the same deviation from the average intercept, the hazard ratio for a one unit increase in individual deviation from the average slope is \( \exp (\alpha_4) \). The code for estimating the joint model with the “shared random effects” parameterization is

\[
\text{jointFit.mlsc3} \leftarrow \text{jointModelBayes(lmeFit.mlsc2.1, Coxfit_fem, timeVar = "Age_50", param = "shared-RE", n.iter = 30000)}
\]

**Table 7**

Parameter Estimates and 95% Credible Intervals for the “Shared Random Effects” Association

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Value</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoker</td>
<td>0.410</td>
<td>0.291</td>
<td>0.536</td>
</tr>
<tr>
<td>AgeStart</td>
<td>-0.048</td>
<td>-0.056</td>
<td>-0.041</td>
</tr>
<tr>
<td>Assocc:(Intercept)</td>
<td>0.004</td>
<td>-0.006</td>
<td>0.014</td>
</tr>
<tr>
<td>Assocc:Age</td>
<td>-1.641</td>
<td>-1.884</td>
<td>-1.397</td>
</tr>
<tr>
<td>tauBs</td>
<td>7.426</td>
<td>2.792</td>
<td>16.424</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>8.330</td>
<td>6.050</td>
<td>10.515</td>
</tr>
<tr>
<td>Age</td>
<td>0.192</td>
<td>0.079</td>
<td>0.308</td>
</tr>
<tr>
<td>AgeStart</td>
<td>-0.046</td>
<td>-0.081</td>
<td>-0.009</td>
</tr>
<tr>
<td>I(Age)</td>
<td>-0.007</td>
<td>-0.007</td>
<td>-0.007</td>
</tr>
<tr>
<td>Age:AgeStart</td>
<td>-0.005</td>
<td>-0.007</td>
<td>-0.003</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.957</td>
<td>0.924</td>
<td>0.993</td>
</tr>
<tr>
<td>D[1, 1]</td>
<td>31.438</td>
<td>29.167</td>
<td>33.827</td>
</tr>
<tr>
<td>D[2, 1]</td>
<td>-0.322</td>
<td>-0.411</td>
<td>-0.224</td>
</tr>
<tr>
<td>D[2, 2]</td>
<td>0.080</td>
<td>0.074</td>
<td>0.085</td>
</tr>
</tbody>
</table>

Note. D[i, j] = ij-element of the covariance matrix of the random effects; tauBs = a hyperparameter relative to the distribution of the baseline hazard function \( h_0(t) \), tauBs is typically not interpreted.

The results are displayed in Table 7. Assocc:(Intercept) represents \( \alpha_3 \) and Assocc:Age represents \( \alpha_4 \) in Equation 14. Subject-specific deviations from the sample
average intercept are not predictive of the risk of death, whereas subject-specific deviations from the sample average slope are. To interpret this latter effect properly, rather than relating the reduction in mortality risk to a one-unit change, we relate it to a one-SD change, where \( D[2, 2] = 0.080 \). This risk reduction is calculated as 
\[
1 - \exp (\alpha_4 \times \text{std}(b_1)) = 1 - \exp (-1.641 \times \sqrt{0.080}) = 0.371.
\]
Thus, for a 1SD decrease in slope, the percent increase in mortality risk is 37.1%. Participants with greater cognitive decline have greater odds of dying than those declining less. Note that with the “shared random effects” association discussed here, the association between the survival and the PS measure is parametrized in terms of subject specific deviation from the overall linear slope of \( \text{Age} \). This is not the same meaning as the “current value plus slope” association in Equation 6, where the association between the survival and the PS measure is parametrized in terms of the instantaneous slope of the PS trajectory (in other words, the current slope represents the velocity: how quickly the PS measures changes as \( \text{Age} \) increases).

**Joint model with interaction** — We have illustrated the three association structures discussed previously in the joint models. Before reaching a final substantive conclusion about our data set, we would like to illustrate an additional extension of the association function \( f(\cdot) \) in Equation 3. We illustrate interactions of predictors with the current true value \( \mu_i(t) \) and/or current rate of change \( \mu'_i(t) \) terms.

We illustrate this by extending the model `jointFit.mslc1` to include an interaction term between the current true value of processing speed \( \mu_i(t) \) and the smoker status. The corresponding time-to-event submodel can be written as

\[
h_i(t) = h_0(t) \exp (\beta_1 \text{Smoker}_i + \beta_2 \text{AgeStart}_i + \alpha_1 \mu_i(t) + \alpha_5 (\mu_i(t) \times \text{Smoker}_i)),
\]

where the interaction coefficient \( \alpha_5 \) expresses the difference in mortality prediction of \( \mu_i(t) \) between smokers (\( \text{Smoker} = 1 \)) compared to non-smokers (\( \text{Smoker} = 0 \)).

The interaction is specified within the `JMbayes` package using the `transFun` argument, which takes as its value a generic function `tf1` that, in turns, takes arguments \( x \) for the longitudinal parameter of interest and `data` for the predictor with which the longitudinal parameter interacts. This `tf1` function has to be saved under its generic name (`tf1`) and defined in the workspace (see Rizopoulos, 2016, for more complex examples, such as multiple transformation functions applied simultaneously to multiple longitudinal parameters). The R code allowing one to write the generic function `tf1` related to the submodel with interaction in Equation 15 is

```r
tf1<-function (x, data) cbind(x, "Smoker" = x * (data$Smoker=='TRUE'))
jointFit.mlsc4 <- update(jointFit.mlsc1, transFun = tf1)
```
In this code, `jointFit.mlsc4` looks for arguments `x` and `data` in the previously estimated joint model. Thus, `jointFit.mlsc4` calls `jointFit.mlsc1` (the joint model with the current value parametrization) and the function `tf1`. `x` is the association structure used in `jointFit.mlsc1` (here the current value \( \mu_i(t) \)) and `data` are the MLSC_ID_fem2 data used in the estimation of `jointFit.mlsc1`. The function \{\texttt{cbind(x, "Smoker" = x * (data$Smoker=='TRUE'))}\} defines the columns of the design matrix \( X_i(t) \) relative to the association structure (see Equation 1). Here \{\texttt{cbind(x, "Smoker" = x * (data$Smoker=='TRUE'))}\} means that we will first estimate a simple effect of \( \mu_i(t) \) (\( x \)), and then an interaction between the current value \( \mu_i(t) \) and the factor Smoker ("Smoker" = \( x \) * (data$Smoker =='TRUE'))). The `tf1` function has to be saved under its generic name (`tf1.R`) in a separate R file and defined in the workspace. This function has to be sourced with the following command:

```r
source(tf1.R)
```

### Table 8

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Value</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoker</td>
<td>0.498</td>
<td>0.306</td>
<td>0.667</td>
</tr>
<tr>
<td>AgeStart</td>
<td>-0.049</td>
<td>-0.057</td>
<td>-0.043</td>
</tr>
<tr>
<td>Assoct</td>
<td>-0.028</td>
<td>-0.034</td>
<td>-0.021</td>
</tr>
<tr>
<td>Assoct:Smoker</td>
<td>0.007</td>
<td>-0.007</td>
<td>0.020</td>
</tr>
<tr>
<td>( \text{tauBs} )</td>
<td>6.989</td>
<td>2.420</td>
<td>18.851</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>8.746</td>
<td>6.453</td>
<td>11.048</td>
</tr>
<tr>
<td>Age</td>
<td>0.191</td>
<td>0.079</td>
<td>0.300</td>
</tr>
<tr>
<td>AgeStart</td>
<td>-0.054</td>
<td>-0.090</td>
<td>-0.016</td>
</tr>
<tr>
<td>( I(Age^2) )</td>
<td>-0.007</td>
<td>-0.007</td>
<td>-0.006</td>
</tr>
<tr>
<td>Age:AgeStart</td>
<td>-0.005</td>
<td>-0.006</td>
<td>-0.003</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.938</td>
<td>0.906</td>
<td>0.968</td>
</tr>
<tr>
<td>( D[1, 1] )</td>
<td>32.501</td>
<td>30.020</td>
<td>35.273</td>
</tr>
<tr>
<td>( D[2, 1] )</td>
<td>-0.346</td>
<td>-0.453</td>
<td>-0.241</td>
</tr>
<tr>
<td>( D[2, 2] )</td>
<td>0.079</td>
<td>0.073</td>
<td>0.084</td>
</tr>
</tbody>
</table>

**Note.** \( D[i, j] \) = \( ij \)-element of the covariance matrix of the random effects; \( \text{tauBs} \) = a hyperparameter relative to the distribution of the baseline hazard function \( h_0(t) \), \( \text{tauBs} \) is typically not interpreted.

The results are presented in Table 8. The coefficient Smoker means that being smoker (versus non-smoker) increases by 64.54% (\( \exp (0.498) - 1 \) x 100) the risk of dying. Assoct (\( \alpha_i \) in Equation 15) represents the association between the current true value
of processing speed and the mortality risk and \textit{Assoct:Smoker} (\(\alpha_5\) in \textit{Equation 15}) represents the interaction effect between the current true value of processing speed and the factor \textit{Smoker} on the mortality risk. These results indicate that the current value of processing speed \(\mu_i(t)\) is negatively associated with the risk of death (\textit{Assoct} = -0.028). Going back to the previous interpretation of the \textit{Assoct} effect of 3.684 units over a 10-year period, rather than a one-unit change (cf. \textit{Figure 4}), we obtain a 9.8% (calculated as \((1 - \exp (-0.028 \times 3.684)) \times 100\)) increase in mortality risk. Note that this effect is very similar to the associative effect previously estimated for the “current value” parameterization, and remains different from zero. The interaction effect \textit{Assoct:Smoker} is unimportant (95% credible interval contains 0), meaning that the effect of the current true value of processing speed on the mortality risk does not differ between smokers and non-smokers.

**Model comparison** — We can finally proceed to compare statistically the various association structures of the estimated joint models. To do so, we rely on the Deviance Information Criterion (DIC; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002), with smaller values indicating better model adjustments to the data. The R code allowing one to compute the joint models’ comparisons is

\[
\text{anova(jointFit.mlsc1, jointFit.mlsc2, jointFit.mlsc3, jointFit.mlsc4)}
\]

and the results of these comparisons can be found in \textbf{Table 9}.

\begin{table}
\begin{center}
\begin{tabular}{lllll}
Model & \(df\) & LPML & DIC & pD \\
\hline
jointFit.mlsc1 & 5591 & -26425.21 & 51793.14 & 5199.108 \\
jointFit.mlsc2 & 5592 & -26333.20 & 51606.23 & 5186.702 \\
jointFit.mlsc3 & 5592 & -26350.51 & 51791.15 & 5229.485 \\
jointFit.mlsc4 & 5592 & -26357.66 & 51773.84 & 5205.650 \\
\end{tabular}
\end{center}
\begin{flushleft}
Note. LPML = log-pseudo-marginal-likelihood value; DIC = deviance information criterion; pD = effective number of model parameters.
\end{flushleft}
\end{table}

Based on the DIC values, the evidence is strong to prefer the \textit{jointFit.mlsc2} model, which includes the current true value and the current rate of change of processing speed as predictors of the risk of death. Recall that within this model, the effect of the current true value in speed on the hazard of dying was not different from zero, whereas the association between the rate of change in speed and survival was strong. More precisely, individuals experiencing a greater rate of change, equivalent to what is
expected over a 10-year interval, have a 19.3% greater mortality risk than those with shallower rates of change. Thus, we can conclude that cognitive performance per se, assessed via perceptual speed, does not relate strongly to mortality. Rather, it is the rate of change in cognitive performance that is predictive of survival, as hypothesized by the terminal decline and terminal drop hypotheses.

**Discussion**

In this tutorial, we presented a didactic overview of the joint longitudinal and time-to-event modeling framework, presented a comparison of various R packages allowing this type of analysis, and illustrated how to use the most comprehensive of these R packages, JMbayes, to estimate a series of simple joint models. We illustrated joint models with real data from a psychological study on adult cognitive development and provided guidelines to choose among a series of theoretically motivated alternative specifications of joint models.

Due to space constraints and to limit complexity, we have not addressed other advanced features of the JMbayes package. These include functionalities for dynamic predictions, where survival probabilities can be estimated for an individual that was not part of the estimation sample, and for less common association structures. For instance, the package allows the “lagged effects” parametrization, where the hazard of experiencing the event at time \( t \) is associated with the level of the longitudinal measure at a previous time point \( t - p \), hence with a \( p \)-lag. For example, we might want to study how cognitive development during childhood might affect mortality later on in life. Thus, the longitudinal measures representing cognitive development during childhood at time \( t - p \) could be related to mortality at time \( t \), with an important (probably around 50-so years) lag \( p \). An other association structure implemented in JMbayes is the “cumulative effects” parametrization, whereby the hazard of experiencing the event at time \( t \) is associated with the whole information relative to the trajectory of the longitudinal measure up to time \( t \). That is, rather than relating the value or slope on the true trajectory to the event, the whole surface under the trajectory constitutes the predictive information. For example, we might want to relate how learning characteristics during a critical period relate to mortality. In this case, performance during a given learning trial might be less informative or mortality than the entire learning history, as represented by the area under the learning curve.

Recently, JMbayes was extended to support simultaneous modeling of multiple longitudinal outcomes (using the function `mvJointModelBayes(·)`). This multiple parametrization may be useful when several longitudinal measures are recorded and may jointly impact the risk of death. For instance if several measure of cognitive development are recorded (as processing speed, fluid intelligence and crystallized intelligence in Rabbitt et al., 2004), they may jointly impact the risk of death. A recent extension also
includes a very useful additional tool for joint modeling: the time-varying association parameters (Andrinopoulou, Eilers, Takkenberg, & Rizopoulos, 2018). Recalling Equation 4, the association parameter between the longitudinal and survival outcomes is allowed to vary in time and therefore can be written $\alpha_1(t)$. We encourage interested readers to explore also the developments of other R packages and other software allowing for joint model estimation.

The primary objective of this tutorial was to present joint modeling of longitudinal and time-to-event data using the JMbayes package. Rather than reviewing all the technical details of the joint modeling methodology, we aimed instead at providing applied researchers in behavioral sciences a didactic and understandable tutorial explaining what this powerful methodology can offer in substantive terms and how to apply it to one’s own research data. We hope this information will encourage adoption of the joint longitudinal-time-to-event modeling framework in research domains where it is currently underutilized and that should benefit from it.

**Funding:** The authors have no funding to report.

**Acknowledgments:** The authors gratefully acknowledge Lifebrain, Grant number 732592 - H2020-SC1-2016-2017/ H2020-SC1-2016-RTD.

**Competing Interests:** The authors have declared that no competing interests exist.

**Supplementary Materials**

For this article the following supplementary materials are available (Cekic, Aichele, Brandmaier, Köhncke & Ghisletta, 2021):

- R codes not included in the main manuscript.

**Index of Supplementary Materials**


**References**


https://doi.org/10.18637/jss.v081.i03


### Appendix

#### Table A1

*Main Functionalities of R Package JM*

<table>
<thead>
<tr>
<th>Property</th>
<th>JM (Rizopoulos, 2010, 2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation method</td>
<td>Frequentist approach with maximum likelihood estimation using an EM algorithm</td>
</tr>
<tr>
<td>Association structure</td>
<td>Current value, current value and slope, shared random effects, lagged effects, cumulative effects and free-form association structures implemented</td>
</tr>
<tr>
<td>Longitudinal submodel</td>
<td>Longitudinal submodel fitted with the <code>nlme(·)</code> function from the <code>nlme</code> package. Allows a univariate normally distributed response and non linear effects of covariates fitted with spline</td>
</tr>
<tr>
<td>Time-to-event submodel</td>
<td>Time-to-event submodel fitted with the <code>coxph(·)</code> function from the <code>survival</code> package. A number of relative risk and accelerated failure time survival model options are available, including Weibull, piecewise PH, Cox PH, and PH with a spline-approximated baseline risk function. A single time-to-event outcome or competing risks are feasible. The time-to-event submodel supported the stratification procedure</td>
</tr>
<tr>
<td>Model comparison</td>
<td>Joint models comparison can be performed with a Likelihood Ratio Test (LRT) implemented with the <code>anova(·)</code> function</td>
</tr>
<tr>
<td>Post-fit analysis</td>
<td>Various post-fit functions including goodness-of-fit analyses, plots, predicted trajectories, individual dynamic prediction of the event and predictive accuracy assessment are available</td>
</tr>
</tbody>
</table>
Table A2

*Main Functionalities of R Package JMbayes*

<table>
<thead>
<tr>
<th>Property</th>
<th>JMbayes <em>(Rizopoulos, 2016)</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation method</td>
<td>Bayesian approach using a MCMC algorithm. A long chain methodology is implemented. Multiple chains are also feasible with a little programming</td>
</tr>
<tr>
<td>Association structure</td>
<td>Current value, current value and slope, shared random effects, lagged effects, cumulative effects and free-form association structure are implemented</td>
</tr>
<tr>
<td>Longitudinal submodel</td>
<td>Continuous and categorical longitudinal outcomes are allowed. Longitudinal submodel fitted with the <em>nlme</em> (·) function from the <em>nlme</em> package. Allows one to fit non linear effects of covariates with spline. A very recent extension of the package allows one to fit multivariate normally distributed continuous and categorical longitudinal outcomes jointly with the time-to-event variable of interest</td>
</tr>
<tr>
<td>Time-to-event submodel</td>
<td>Time-to-event submodel fitted with the <em>coxph</em> (·) function from the <em>survival</em> package. A number of relative risk and accelerated failure time survival model options are available, including Weibull, piecewise PH, Cox PH, and PH with a spline-approximated baseline risk function. The time-to-event submodel supported the stratification procedure</td>
</tr>
<tr>
<td>Models comparison</td>
<td>Joint models comparison can be performed with the Deviance Information Criterion (DIC) available with the <em>anova</em> (·) function</td>
</tr>
<tr>
<td>Post-fit analysis</td>
<td>Various post-fit tools for model diagnostic checks are available as diagnostic plots. Functionalities are available for computing dynamic predictions for both the longitudinal and time-to-event outcomes and assessment of model accuracy in terms of discrimination and calibration</td>
</tr>
<tr>
<td>Extensions</td>
<td>Joint modeling for multivariate longitudinal outcomes and for time-varying association structures using P-splines are two very recent extensions implemented within the JMbayes package</td>
</tr>
</tbody>
</table>
### Table A3

**Main Functionalities of R Packages joineR and joineRML**

<table>
<thead>
<tr>
<th>Property</th>
<th>joineR (Philipson et al., 2017; Williamson, Kolamunnage-Dona, Philipson, &amp; Marson, 2008)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation method</td>
<td>Frequentist approach with maximum likelihood estimation using an EM algorithm</td>
</tr>
<tr>
<td>Association structure</td>
<td>The implemented associations structures are based on an extended version of the “shared random effects” model proposed by Wulfsohn and Tsiatis (Wulfsohn &amp; Tsiatis, 1997)</td>
</tr>
<tr>
<td>Longitudinal submodel</td>
<td>Longitudinal submodel is univariate and fitted with a linear mixed-effects model (splines and non normal responses are not feasible)</td>
</tr>
<tr>
<td>Time-to-event submodel</td>
<td>Time-to-event submodel is a Cox PH model with log-Gaussian frailty</td>
</tr>
<tr>
<td>Models comparison</td>
<td>Models comparison can be performed with Likelihood methods</td>
</tr>
<tr>
<td>Post-fit analysis</td>
<td>Exact standard errors intervals can be obtained with implemented bootstrap methodology</td>
</tr>
<tr>
<td>Extensions</td>
<td>The joineRML package (Hickey, Philipson, Jorgensen, &amp; Kolamunnage-Dona, 2016, 2018) is a recent extension of the joineR which allows one to fit multivariate linear longitudinal data with a correlated time-to-event using Bayesian Monte Carlo EM algorithm. joineRML also includes functions to derive dynamic predictions.</td>
</tr>
</tbody>
</table>

### Table A4

**Main Functionalities of R Package lcmm**

<table>
<thead>
<tr>
<th>Property</th>
<th>lcmm (Proust-Lima, Philipps, Diakite, &amp; Liquet, 2017)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation method</td>
<td>Frequentist approach based on maximum likelihood estimation using a modified Marquardt algorithm</td>
</tr>
<tr>
<td>Association structure</td>
<td>The associations structures available are based on joint latent class models assumptions (joint models that consider homogeneous latent subgroup processing speed of individuals sharing the same longitudinal trajectory and risk of event)</td>
</tr>
<tr>
<td>Longitudinal submodel</td>
<td>The linear model can include an univariate Gaussian outcome, an univariate curvilinear outcome, an univariate ordinal outcome and curvilinear multivariate outcomes</td>
</tr>
<tr>
<td>Models comparison</td>
<td>Models comparison can be performed with likelihood methods</td>
</tr>
<tr>
<td>Post-fit analysis</td>
<td>Various post-fit functions with goodness-of-fit analyses, classification, plots, predicted trajectories, individual dynamic prediction of the event and predictive accuracy assessment are available</td>
</tr>
</tbody>
</table>
### Table A5

**Main Functionalities of R Package frailtypack**

<table>
<thead>
<tr>
<th>Property</th>
<th>frailtypack (Król et al., 2017; Rondeau, Mazroui, &amp; Gonzalez, 2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation method</td>
<td>Frequentist maximum likelihood approach using either a parametric or semiparametric approach on the penalized likelihood for estimation of the hazard functions</td>
</tr>
<tr>
<td>Longitudinal submodel and time-to-event submodel</td>
<td>Several type of joint models are implemented. In particular, joint models for recurrent events and a terminal event, for two time-to-event outcomes for clustered data, for two types of recurrent events and a terminal event, for a longitudinal biomarker and a terminal event and joint models for a longitudinal biomarker, recurrent events and a terminal event</td>
</tr>
<tr>
<td>Models comparison</td>
<td>Two criteria for assessing model's predictive accuracy are implemented and can be used for models comparison</td>
</tr>
<tr>
<td>Post-fit analysis</td>
<td>Each model function allows to evaluate goodness-of-fit analyses and provides plots of baseline hazard functions. Individual dynamic predictions of the terminal event and evaluation of predictive accuracy are also implemented</td>
</tr>
</tbody>
</table>

### Table A6

**Main Functionalities of R Package rstanarm**

<table>
<thead>
<tr>
<th>Property</th>
<th>rstanarm (Goodrich, Gabry, Ali, &amp; Brilleman, 2018)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation method</td>
<td>Bayesian approach using a MCMC algorithm with the function <code>stan_jm(·)</code>. A multiple chain methodology is implemented (long chain is also feasible). Very clear and user friendly implementation</td>
</tr>
<tr>
<td>Association structure</td>
<td>Current value, current value and slope, shared random effects, lagged effects, cumulative effects and interaction effect associations structures are implemented</td>
</tr>
<tr>
<td>Longitudinal submodel</td>
<td>Generalized linear mixed-effects model (the response has to belong to an exponential family distribution). The longitudinal part can be multivariate, continuous and/or categorical. Linear slope, cubic splines and polynomial terms are allowed</td>
</tr>
<tr>
<td>Time-to-event submodel</td>
<td>The baseline hazard can be specified parametrically or non parametrically. A stratification procedure is not allowed</td>
</tr>
<tr>
<td>Models comparison</td>
<td>There is currently no models comparison procedure implemented</td>
</tr>
<tr>
<td>Post-fit analysis</td>
<td>Several post-fit functions for dynamic predictions of the terminal event and longitudinal trajectories and as well as visualisation functions are available</td>
</tr>
</tbody>
</table>
Table A7

*Main functionalities of R Package `bamlss`*

<table>
<thead>
<tr>
<th>Property</th>
<th><code>bamlss (Umlauf, Klein, &amp; Zeileis, 2018)</code></th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation method</td>
<td>Bayesian method</td>
</tr>
<tr>
<td>Association structure</td>
<td>Flexible additive joint models are implemented</td>
</tr>
<tr>
<td>Longitudinal submodel</td>
<td>Univariate continuous longitudinal outcome is allowed an</td>
</tr>
<tr>
<td>Time-to-event submodel</td>
<td>A single time-to-event outcome is allowed</td>
</tr>
</tbody>
</table>