

Bivariate joint models for survival and change of cognitive function

Shengning Pan  and Ardo van den Hout

Statistical Methods in Medical Research

2023, Vol. 32(3) 474–492

© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09622802221146307

journals.sagepub.com/home/smm



Abstract

Changes in cognitive function over time are of interest in ageing research. A joint model is constructed to investigate. Generally, cognitive function is measured through more than one test, and the test scores are integers. The aim is to investigate two test scores and use an extension of a bivariate binomial distribution to define a new joint model. This bivariate distribution model the correlation between the two test scores. To deal with attrition due to death, the Weibull hazard model and the Gompertz hazard model are used. A shared random-effects model is constructed, and the random effects are assumed to follow a bivariate normal distribution. It is shown how to incorporate random effects that link the bivariate longitudinal model and the survival model. The joint model is applied to the English Longitudinal Study of Ageing data.

Keywords

Joint model, bivariate binomial distribution, cognitive function, survival analysis, shared random-effects model

1 Introduction

In ageing research, data from various tests of cognitive functioning are often collected as longitudinal data, i.e. the same sample is tracked at different points in time. Generally, longitudinal data on ageing research are multi-dimensional data that contain observations for multiple phenomena obtained over multiple time periods for the same individual.¹ For instance, the English Longitudinal Study of Ageing (ELSA) contains several cognitive tests relating to literacy, numeracy, memory and information processing.² Individuals are tested every few years and the test scores are recorded; these test scores are in the form of non-negative integers. An important research topic in ageing research is the change in cognitive function over time. In our model, we present an extension of the binomial model to measure the change in non-negative cognitive test scores.

Cognitive function, also known as cognitive ability, is the individual's ability to process information, mainly with regard to learning and problem-solving. Researchers are committed to obtaining scores that reflect the level of cognitive function of individuals through different tests.

In ageing research, it is important to investigate the changes in the individual's cognitive function. The relationship between changes in cognitive function and the risk of death is also of interest.³ Researchers provide advice on whether the elderly need care by analyzing the relationship between cognitive function decline and ageing.⁴ For example, an individual whose cognitive function drops suddenly during a period might suffer from a disease such as Alzheimer's disease that affects cognitive abilities or even threatens life, and he/she needs professional medical care. The work in Laird and Ware,⁵ Cox⁶ and Van Den Hout and Muniz-Terrera⁷ model longitudinal data and time-to-event data separately, however, if the two response processes are correlated this may lead to biased effect size estimates.⁸ Therefore, it is

Department of Statistical Science, University College, London, UK

Corresponding author:

Shengning Pan, Department of Statistical Science, University College London, United Kingdom.

Email: shengning.pan.18@ucl.ac.uk

reasonable to use a joint model to measure both processes simultaneously. The use of joint models is not limited to the ageing research. Previous research has shown that joint modelling can improve the efficiency of statistical inference and reduce bias, and it can provide significant benefits when designing experiments.^{9,10}

Research into the joint modelling of longitudinal and time-to-event data has received considerable attention over the past three decades.^{3,11} Faucett and Thomas, and Wulfsohn and Tsiatis wrote seminal articles in this field in 1996 and 1997 respectively, and after that extensions to joint models have been proposed in the literature.¹² In the early stages of research on joint models, research on joint models has focused on single longitudinal responses and single time-event responses, i.e. univariate joint model. This approach has been widely used in clinical research¹³ and there are a number of mainstream software packages available for calculating related problems.^{14–16} However, as we mentioned above, longitudinal and time-to-event data, especially the longitudinal part, are usually multidimensional and contain observations of multiple phenomena obtained by the same person over multiple time periods. While univariate joint models allow us to determine the relationship between a survival outcome and a single longitudinal outcome, more frequently, multiple biomarkers are associated with the event of interest.¹² Multivariate joint models allow us to consider more information simultaneously, thus enabling us to better understand the dynamic complexity of changes in disease or biological indicators. In recent years, a number of researchers have devoted their research to the multivariate joint model. For example^{17–19} are multivariate joint model investigations with continuous response variables. There are also many articles focusing on other data types, e.g. article²⁰ considers multivariate binary data, the model in Rizopoulos and Ghosh²¹ combines continuous response variables with binary counterparts and article by Wang et al.²² uses ordinal data. Rue et al.²³ build a joint multivariate model related to discrete data. Although the responses in Rue et al.²³ are discrete variables, the longitudinal data model in the article still uses the normal distribution model and the beta model. Many software packages also provide relevant functionality for multivariate joint models. Examples include *merlin* in *Stata*²⁴ and the *JMbayes* package in *R*.²⁵ In this paper, we construct a bivariate discrete joint model to address the potential estimation bias caused by separate modelling and to avoid the lack of information caused by dropping longitudinal responses when using a univariate joint model. This joint model can take into account discrete data.

A joint model for longitudinal responses describes the change of responses using a model for longitudinal data and describes the risk of the event using a survival model.²⁶ The principle of joint models is to link these two models using a common latent variable. This common latent variable captures the association between the processes. Based on this principle, two joint models have been proposed: the shared random effects model¹¹ and the latent class model.²⁷ Articles Rouanet et al.,²⁸ Proust-Lima et al.²⁶ and Proust-Lima et al.²⁹ used latent class models to construct multivariate joint models. The limitation of the latent class model is that the classification may not be interpretable because we do not know what the different classes represent; moreover, we can only calculate the probability that an individual belongs to a certain class and cannot assign him or her to a specific class with certainty. To avoid these limitations, we model the latent variable as a random effect. Conditional on random effects we assume that the longitudinal outcome is independent of the time-to-event outcome.¹⁴ Also, we assume that the repeated measurements in the longitudinal outcome are independent of each other given the random effects.

A multivariate shared random effects model was introduced by Rizopoulos in Rizopoulos.¹⁴ In this model, the association between longitudinal variables is captured by the random effects. Our bivariate joint model uses an extended bivariate binomial distribution³⁰ in the longitudinal part of the model. This distribution includes a parameter that represents the correlation coefficient between the two longitudinal outcomes. The model allows us to avoid using random effects to capture all the correlations of the joint model.

In this paper, we will investigate the bivariate responses changing over time, in which the responses are different measurements representing cognitive function for individuals. The shared random effects can take into account unobservable individual-specific characteristics.¹⁴ Since the cognitive function is usually measured by non-negative integers, the bivariate extension of the binomial distribution proposed by Altham and Hankin³⁰ is used for the longitudinal model. This model not only measures whether the model is over- or underdispersed compared to the standard binomial distribution but also measures the association between the two longitudinal outcomes. For the survival part, the Weibull hazard model and the Gompertz hazard model are used. In univariate joint models, the link functions which contains in both the longitudinal model and the survival model are well defined.³¹ However, for the bivariate joint model, we do not know beforehand which link will provide the best result. We will therefore discuss several different link functions.

We apply the models to analyze the ELSA data. There are two non-negative longitudinal outcomes in the ELSA data, and they are related to verbal learning and recall. Individuals were asked to learn ten words and recall them at two different time points (immediate and later), but in the same interview; see Van den Hout and Muniz-Terrera.⁷ In the ELSA data, the event time for the joint model is the time of death.

The ‘Models section provides a brief overview of the models in this paper: the extension of the bivariate binomial distribution proposed by Altham and Hankin, the Weibull hazard model and the Gompertz survival model. The link between the

longitudinal model and the survival model is also discussed in this part. The ‘Maximum likelihood function given left truncation section constructs the marginal log-likelihood function for the joint model, discusses the left truncation and proposes the method for computation. In the ‘Simulation section, we use a simulation study to validate the accuracy of the joint model parameter estimates. We analyze the ELSA data through fitting the shared random-effects model in the ‘Application section. The conclusion of this paper and possible extensions options for the joint model are discussed in the ‘Conclusion section.

2 Models

The joint model is given by

$$p(\mathbf{y}, \mathbf{t}, t_{last}) = \int p(\mathbf{y}|\mathbf{t}, \mathbf{b})p(t_{last}|\mathbf{b})p(\mathbf{b})d\mathbf{b} \quad (1)$$

where \mathbf{y} represents the longitudinal outcomes, \mathbf{t} represents the corresponding time points, t_{last} is the last recorded time and \mathbf{b} represent the random effects. Distribution $p(\mathbf{y}|\mathbf{t}, \mathbf{b})$ is the longitudinal model, $p(t_{last}|\mathbf{b})$ is the survival model and \mathbf{b} represent the random effects.

The first part of this section introduces the bivariate binomial distribution used for the longitudinal model. The second part proposes the hazard survival model. The third part defines the link between the longitudinal model and the survival model, which consists of parameters common to both models and the random effects. For what follows, assume that the random effects are given by vector $\mathbf{b} \in \mathbb{R}^p$, where p is determined by the model definition. We will use $p(\mathbf{b})$ to represent the distribution of random effects.

2.1 Longitudinal model

Generally, the scores for cognitive function are discrete non-negative integers. For example, the Mini-Mental State Examination (MMSE) test is a 30-point questionnaire, widely used extensively in clinical and research settings to measure cognitive impairment. A higher score represents better cognitive function. We model the probability of success (answering the question correctly and earning one point). For the MMSE, this would imply the assumption that the 30 questions are independent of each other and that the distribution of the total-score follows the binomial distribution. It is worth to notice that even if the probability of success is different, the total-score distribution still follows the binomial distribution.³²

Let Y denote the response and p denote the probability of success. Parameter m represents the number of trials. Altham³³ introduced the following extension of the binomial distribution.

$$p(Y = y) = \frac{\binom{m}{y} p^y (1-p)^{(m-y)} \theta^{y(m-y)}}{C_{uni}} \quad (2)$$

for $y \in \{0, 1, \dots, m\}$, where $0 < p < 1$, $\theta > 0$, and

$$C_{uni} = \sum_{y=0}^m \binom{m}{y} p^y (1-p)^{(m-y)} \theta^{y(m-y)} \quad (3)$$

When $\theta > 1$, distribution (2) is under-dispersed compared to the standard binomial distribution; when $\theta < 1$, the distribution is over-dispersed relative to the standard distribution.

In this paper, we use the bivariate extension of (2)³⁰ to analyze the longitudinal data. Let Y_1 and Y_2 define the bivariate responses.

$$p(Y_1 = y^{(1)}, Y_2 = y^{(2)}) = \frac{g(Y_1 = y^{(1)})g(Y_2 = y^{(2)})\phi^{y^{(1)}y^{(2)}}}{C_{biv}} \quad (4)$$

for $y^{(j)} \in \{0, 1, \dots, m_j\}$, $j = 1, 2$.

$$g(Y_j = y^{(j)}) = \binom{m_j}{y^{(j)}} p_{Y_j}^{y^{(j)}} (1 - p_{Y_j})^{(m_j - y^{(j)})} \theta_{Y_j}^{y^{(j)}(m_j - y^{(j)})} \quad (5)$$

where p_{Y_j} denote the probability of success for $j = 1, 2$; $0 < p_{Y_1}, p_{Y_2} < 1$, $\theta_{Y_j}, \phi > 0$. The parameter ϕ is used to measure the correlation between two responses. When $0 < \phi < 1$, there is a negative correlation between two responses; when ϕ is

greater than 1, there is a positive correlation between responses. The denominator in Equation (4) is:

$$C_{biv} = \sum_{y^{(1)}=0}^{m_1} \sum_{y^{(2)}=0}^{m_2} g(Y_1 = y^{(1)})g(Y_2 = y^{(2)})\phi^{y^{(1)}y^{(2)}}$$

Integers m_1 and m_2 for longitudinal responses $y^{(1)}$ and $y^{(2)}$ can be defined separately.

Assume that the random variable $b = (b_{10}, b_{11}, b_{20}, b_{21})^T$ includes the random intercepts (b_{10} and b_{20}) and the random slopes (b_{11} and b_{21}). Parameters b_{j0} and b_{j1} denote the corresponding random effects for $Y_j, j = 1, 2$. Let t represent time or age.

To investigate the relationship between time and the response variables, we use the random-effects logistic regression model:

$$p_{Y_j} = \frac{\exp\left((\eta_0^{(j)} + b_{j0}) + (\eta_1^{(j)} + b_{j1})t\right)}{1 + \exp\left((\eta_0^{(j)} + b_{j0}) + (\eta_1^{(j)} + b_{j1})t\right)} \quad (6)$$

where η s are the fixed effects in the model. We could add other covariates to model (6), such as sex and education level, to investigate the effect of these covariates on an individual's cognitive function. We use x to denote any covariate, and the extension of model (6) containing a covariate could be written as:

$$p_{Y_j} = \frac{\exp\left((\eta_0^{(j)} + b_{j0}) + (\eta_1^{(j)} + b_{j1})t + \gamma_L^{(j)}x\right)}{1 + \exp\left((\eta_0^{(j)} + b_{j0}) + (\eta_1^{(j)} + b_{j1})t + \gamma_L^{(j)}x\right)} \quad (7)$$

the parameter $\gamma_L^{(j)}$ is the coefficient of the corresponding covariate and the corner label L represents the longitudinal model.

In this paper, we will investigate restricted models. We will explain the reasons for using restricted models in the later section. The restricted models are defined by restrictions on random effects, for example:

$$b_{10} = b_{20}, b_{11} = b_{21} \quad (8)$$

where two responses Y_1 and Y_2 share the same random intercept b_{10} and the same random slope b_{11} .

We can also set

$$b_{11} = b_{21} = 0 \quad (9)$$

which means that we only use random intercept for each response. Also, we can assume

$$b_{10} = b_{20} = 0 \quad (10)$$

where responses have different random slope (b_{11} and b_{21}) and do not have corresponding random intercepts.

The reason for choosing this bivariate binominal distribution is that the formula of the probability density function is relatively simple, and it belongs to the exponential family.³⁰ Moreover, this distribution allows Y_1 and Y_2 to have separate choices of m_1 and m_2 .

2.2 Hazard models

We define the hazard models and the corresponding survivor models for analyzing the time-to-event data. The hazard model is a parametric regression model given by

$$h(t) = h_0(t) \exp(\Delta(\alpha, \eta, \mathbf{b}, \gamma_S, t)) \quad (11)$$

where $h_0(t)$ is the baseline hazard function. Function Δ is a function of random effects, additional parameter α and common parameters. We define function Δ as the 'shared part' (link function) in the joint model, and it will be introduced in the next section. Vector $\eta = (\eta_0^{(j)}, \eta_1^{(j)})$ represents the fixed parameters. Similar to equation (7), γ_S is the parameter of the covariate and the corner label S represents the survival model. In the application in the 'Simulation section, time t in the survival model is the last recorded age.

There are many widely used parametric hazard models in survival analysis, such as the exponential hazard model, the Weibull hazard model, the Gompertz hazard model, and the log-logistic distribution. In this paper we use two commonly adopted models to construct the survival part of the joint model: the Weibull hazard model and the Gompertz hazard model.

The Weibull model is a survival model with two positive parameters and it is a generalization of the exponential model. We denote the Weibull model $Weibull(\lambda, \tau)$. The shape of the function can be changed by varying the value of τ . In particular, when τ equal to 1 we obtain an exponential distribution. Xian Liu has mentioned that of all the families of parametric time distributions, the Weibull model is probably the most widely used parametric function in survival analysis due to its simplicity and flexibility,³⁴ therefore we have chosen it as the first survival model used in this paper. The Gompertz model is the most widely used survival function to quantitatively describe human mortality and survival.^{35,34} The numerical estimation for the Gompertz model is computational robust compared with the Weibull model because there is no power function in the Gompertz model. Given that the event of interest is death, we have chosen the Gompertz model as the second survival model. Although in this paper we restrict the survival model to the Weibull model and the Gompertz model, the survival part of our joint model could be easily replaced by other frameworks.

We have the following baseline hazard models:

$$\begin{aligned} \text{Weibull: } h_0(t) &= \lambda \tau t^{\tau-1} \\ \text{Gompertz: } h_0(t) &= \lambda \exp(\xi t) \end{aligned} \quad (12)$$

where $\lambda, \tau > 0$, and the corresponding survivor functions are:

$$\begin{aligned} \text{Weibull: } p(T \geq t) = S(t) &= \exp(-\lambda t^\tau) \\ \text{Gompertz: } p(T \geq t) = S(t) &= \exp\left(-\frac{\lambda}{\xi} \left(\exp(\xi t) - 1\right)\right) \end{aligned} \quad (13)$$

For computational reasons, we use the parametrization $\lambda = \exp(\beta)$ in the joint model to ensure the hazard is positive for any $\beta \in \mathbb{R}$.

The definition for the range of ξ in the Gompertz model varies in the literature. Some researchers define $\xi \in \mathbb{R}$, to ensure that the hazard exponentially increases over time when $\xi > 0$ and decreases when $\xi < 0$.³⁶ However, if $\xi < 0$ and $t \rightarrow \infty$, we have the following property of the survivor function:

$$\lim_{t \rightarrow \infty} S(t) = \exp\left(\frac{\lambda}{\xi}\right) > 0 \quad (14)$$

The above equation shows that when the time is extremely large, the probability of survival is still larger than 0. It means that the event of interest does not occur for a proportion of the population. The events of interest in this paper are death, and the process from health to event occurring is irreversible. Therefore, we follow the definition given by Wienke³⁷ and restrict ξ to be positive. In this case, when $t = 0$ the hazard function equals to λ , and it increases to ∞ when $t = \infty$. The corresponding survivor function equals to 1 when $t = 0$, and it decreases to 0 when t goes up to ∞ .

2.3 Link between the models

In this part, we introduce the link function $\Delta(\alpha, \eta, \mathbf{b}, \gamma_S, t)$ of the model, defined as an expression containing the random effects and the corresponding parameters. Parameters in this part are included in both the longitudinal model and the survival model. Based on the link function, we will make some remarks for the corresponding hazard models. Since in this section, we will focus on the structure of the random intercept and random slope, we will simplify the link function $\Delta(\alpha, \eta, \mathbf{b}, \gamma_S, t)$ to $\Delta(\alpha, \eta, \mathbf{b}, t)$.

The hazard model (11) allows for several specializations. In joint models, $\Delta(\alpha, \eta, \mathbf{b}, t)$ is often specified as $\alpha(\eta_0 + b + (\eta_1 + b)t)$ for the univariate model.³¹ In many articles related to multivariate joint models, the linear predictor is also commonly used as the link function. In our paper, we also use the linear predictor as the link function. However, when investigating the bivariate joint model, we do not know beforehand which kind of link will provide the best result. As we have mentioned before, we will use the restricted model based on equations (8) to (10) and construct the corresponding link function. We use restricted models for two reasons. Firstly, if we use the full model, i.e. models containing equations (6) or (7) for $j = 1, 2$, it would be unclear how to construct the link between the models and how to interpret the impact of the corresponding parameters on the risk of death. Secondly, we encountered challenges in computing high-dimensional integrals. Therefore, we reduced the dimensionality of the integrals, i.e. the number of corresponding random effects, while retaining the study of response-specific random effects (each response variable has a corresponding random effect). We consider a range of definitions of Δ , starting with

$$\Delta(\alpha, \eta, \mathbf{b}, t) = \alpha\left(\eta_0^{(j)} + b_{j0}\right) \quad (15)$$

where $\eta_0^{(j)} + b_{j0}$ is the random intercept for response Y_j , for $j = 1$ or 2 . We can also include the random intercept for both Y_1

and Y_2 :

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \mathbf{b}, t) = \sum_{j=1}^2 \alpha_j (\eta_0^{(j)} + b_{j0}) \quad (16)$$

The advantage of (15) and (16) is that they are computationally easy because the link part is not affected by time. However, comparing with the value of the cognitive function at the start of the time scale, change in cognitive ability over time might have a greater impact on the risk of death or dementia. Therefore, we can include the random slopes in the same way.

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \mathbf{b}, t) = \sum_{j=1}^2 \alpha_j (\eta_1^{(j)} + b_{j1}) t \quad (17)$$

For the above link function containing random slopes, we also consider the restricted version, i.e. setting one of the α_j s to 0

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \mathbf{b}, t) = \alpha (\eta_1^{(j)} + b_{j1}) t \quad (18)$$

Equation (17) takes into account the trajectory of Y_1 and Y_2 changing over time by adding the fixed slope and the random slope. We can also use the random intercept and the random slope at the same time.

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \mathbf{b}, t) = \sum_{j=1}^2 \alpha_{1j} (\eta_0^{(j)} + b_0) + \alpha_{2j} (\eta_1^{(j)} + b_1) t \quad (19)$$

where the random effects are intercept/slope specific, and Y_1, Y_2 share the same random effects. We wish to distinguish the effect of the two response variables on the risk, therefore models with link function (19) will not be included in the application part of the research.

In addition to the link functions above, there are many other common forms of link functions. For example, in many joint models which use a linear mixed model as their longitudinal model, the linear predictor is the conditional expectation of the response variables. Such models make it possible to use the expectation in constructing the link function:

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \mathbf{b}, t) = \alpha_j (\eta_0^{(j)} + b_{j0} + (\eta_1^{(j)} + b_{j1}) t) \quad (20)$$

Equation (20) is a common form of using expectation $E(Y_j)$ when constructing a link function. However, in our distribution, the expectation of the response is not the linear predictor. The expression in our joint model is a function of the range of the response m and the probability of success p_{Y_j} .

When substituting the above Δ expressions into the hazard models, we need to pay attention to the parameters λ, τ and ξ in the corresponding hazard functions.

For the Weibull hazard, when the expression of Δ is equal to (15) or (16). The hazard model can be written as:

$$h(t) = \lambda^* \tau t^{(\tau-1)} \quad (21)$$

where the scale parameter λ is now written as an individual-specific parameter:

$$\lambda^* = \exp(\beta + \alpha(\eta_0^{(j)} + b_{j0})) \quad (22)$$

The Weibull hazard model can also be written as a hazard model with an individual-specific slope parameter:

$$h(t) = \lambda \tau^* t^{(\tau^*-1)} \quad (23)$$

where

$$\tau^* = \exp(\alpha(\eta_0^{(j)} + b_{j0})) \tau \quad (24)$$

In this paper, we choose equation (21) as the corresponding hazard model. When the Δ s include the random slopes (equations (18) and (19)) are substituted into Weibull hazard model, the shape parameter τ becomes a time-dependent parameter, which violates the definition of the standard Weibull model. When we include random slopes in the Weibull model, we have to calculate the log cumulative hazard function.³⁸ In this paper, we will not discuss this case of using the log

cumulative hazard function.

For the Gompertz hazard, we have:

$$h(t) = \lambda^* \exp(\xi^* t) \quad (25)$$

where

$$\begin{aligned} \lambda^* &= \exp(\beta + \alpha(\eta_0^{(j)} + b_{j0})) \\ \xi^* &= \alpha(\eta_1^{(j)} + b_{j1}) + \xi \end{aligned} \quad (26)$$

Both λ^* and ξ^* are individual-specific parameters. The Gompertz hazard model can easily handle the function Δ with random slopes. Because the model conditional on random effects is still a Gompertz hazard model.

2.4 Distribution of random effects

As we mentioned at the beginning of the ‘Models section, we use $p(\mathbf{b})$ to represent the distribution of random effects. In this paper, we use the bivariate normal distribution as the distribution of random effects, i.e. $\mathbf{b} \sim \left(0, \begin{pmatrix} \sigma_{b1}^2 & \rho\sigma_{b1}\sigma_{b2} \\ \rho\sigma_{b1}\sigma_{b2} & \sigma_{b2}^2 \end{pmatrix}\right)$. Parameter ρ represents the correlation between random effects. The parameters σ_{b1} and σ_{b2} denote the standard deviation of the corresponding random effects respectively.

3 Maximum likelihood function given left truncation

Left truncation, also called *delayed entry*, occurs when individuals have been at risk before entering the study.³⁷ For example, when the event of interest is death, individuals who died before the study started will not be included in the study. In ageing research, the main timescale is age. Since the event of interest is death, individuals can be included in the data only if they have not experienced the event before they enter the study. If we do not deal with the left truncation, the estimation is based on the assumption that individuals were not at risk of dying before the start of the study. Therefore, the left truncation needs to be taken into account in the model estimation.

For individual i , $i = 1, \dots, N$, the corresponding longitudinal responses are $\mathbf{y}_i = (\mathbf{y}_i^{(1)}, \mathbf{y}_i^{(2)})$. Response $\mathbf{y}_i^{(j)}$ is a vector $(\mathbf{y}_i^{(j)} = (y_{i1}^{(j)}, \dots, y_{in_i}^{(j)}))$ at age $\mathbf{t}_i = (t_{i1}, \dots, t_{in_i})$, where $j = 1, 2$ is the j th method of measuring cognitive function, n_i is the number of observations for each individual. Let ω represent all the parameters in the joint model except the random effects. Let t_{i1} denote baseline age. The likelihood contribution of individual i conditionally on truncation time t_{i1} is:

$$\begin{aligned} L_i(\omega | \mathbf{y}_i, \mathbf{t}_i, T \geq t_{i1}) &= p(\mathbf{y}_i, \mathbf{t}_i | T \geq t_{i1}, \omega) \\ &= \frac{p(\mathbf{y}_i, \mathbf{t}_i | \omega)}{p(T \geq t_{i1} | \omega)} \end{aligned} \quad (27)$$

where $p(T \geq t_{i1} | \omega)$ is the survivor function evaluated at t_{i1} . For the shared random-effects model, the denominator in (27) can be written as:

$$p(T \geq t_{i1} | \omega) = \int p(T \geq t_{i1} | \mathbf{b}_i, \omega) p(\mathbf{b}_i | \omega) d\mathbf{b}_i$$

Assuming independence between responses given the random effect, the marginal likelihood function is:

$$p(\mathbf{y}_i, \mathbf{t}_i | \omega) = \int p(\mathbf{y}_i | \mathbf{t}_i, \mathbf{b}_i, \omega) p(t_{in_i} | \mathbf{b}_i, \omega) p(\mathbf{b}_i | \omega) d\mathbf{b}_i \quad (28)$$

where t_{in_i} is the last recorded age. Parameter $\delta_i = 0$ means alive at the last observation and $\delta_i = 1$ means death. Distribution $p(\mathbf{y}_i | \mathbf{t}_i, \mathbf{b}_i)$ is defined by the longitudinal model and $p(t_{in_i} | \mathbf{b}_i, \delta_i)$ by the survival model:

$$p(t_{in_i} | \mathbf{b}_i, \delta_i) = h(t_{in_i} | \mathbf{b}_i)^{\delta_i} p(T \geq t_{in_i})$$

We define the random effect $\mathbf{b}_i \in \mathbb{R}^p$ by $\mathbf{b}_i \sim N(0, \Sigma)$, where Σ is a $p \times p$ covariance matrix.

This model offers a flexible way to model the correlation between the value change of responses and the risk of event.³⁹ The advantage of using the normal distribution is that it is one of the most common distributional choices in linear mixed-effects models and that it enables using Gauss-Hermite quadrature in the numerical optimization of the marginal

likelihood. However, even if we can use the Gauss-Hermite integration, this model is still computationally intensive due to the numerical integration.

We code the marginal log-likelihood function in the *R* software.⁴⁰ The corresponding parameters are estimated using the *ucminf* function in package *ucminf*.⁴¹ For the shared random-effects model (28), where the random effect b follows multivariate normal distribution, we use the Gauss-Hermite method with 7 nodes to do numerical integration. The accuracy of the node number selection will be verified in the ‘Simulation section.

It is worth mentioning that the bivariate Gauss-Hermite quadrature can be undertaken by using two univariate normal distributions (see, e.g. Van Den Hout³⁹). The bivariate normal distribution $f(z, x)$ can be expressed as:

$$f_{Z|X}(z|X = x)f_X(x) \quad (29)$$

where $Z \sim N(\mu_Z, \sigma_Z^2)$, $X \sim N(\mu_X, \sigma_X^2)$ and $Z|X \sim N(\mu_Z - \rho(\sigma_Z/\sigma_X)x - \mu_X, \sigma_Z^2(1 - \rho^2))$. Parameter ρ represents the correlation between Z and X .

4 Simulation

We conduct a small simulation study to investigate the parameter estimation for the joint model. We used the ADEMP structure to plan the simulation study.⁴² The ADEMP structure includes *Aims*, *Data-generating mechanisms*, *Methods*, *Estimands*, *Performance measures*.

The main *Aims* of our simulation study are to check the estimation method and to investigate the performance of joint model given various sample sizes.

Date-generating mechanism: We use the joint model with the best results in the ‘Application section as the model for the ‘Simulation section. The mechanism is based on the extension of the bivariate binomial distribution (4) and the Gompertz hazard model (13). As Morris et al.⁴² mentioned, varying the sample size of the simulation dataset is a common approach when changing the data generation mechanism since the performance tends to vary with the sample size. We also refer to the article by Van den Hout and Muniz-Terrera,⁴ which designed a small simulation study in a similar setting. Therefore, to investigate the small sample bias, for the joint model we set the sample size $N = 100, 400, 1000$, respectively.

For the joint model, the logistic regression model is defined by (6) with $b_{11} = b_{21} = 0$, with the corresponding link function (15) for $j = 1$. The random-effects distribution for (b_{10}, b_{20}) is specified by $\sigma_{b_1}, \sigma_{b_2}$ and ρ . This joint model is the best performing model in the application; see the ‘Application section.

The true values of the parameters, i.e. the values used when generating the data, are listed as ‘Value’ in Table 1. It is worth noting that we set α to -0.4 in the model. The parameter α is the relevant parameter of the link function $\Delta(\alpha, \eta, \mathbf{b}, t)$, if α is significantly different from 0, it means that there is an association between the longitudinal model and the survival model, and it is necessary for us to build the joint model.

Estimands: In this simulation study, the estimands of interest are the model parameters over the N_{sim} iterations.

Methods: In the simulation study, to eliminate the effect of left truncation, we set the observation time for all individuals beginning at time zero. If an individual does not drop out, the follow-up ends at 24 years. Follow-up intervals are set to three years to ensure that we have enough information about changes in response variables over time. When the follow-up interval was fixed at once every three years, it means that there are 9 observations for individuals who do not drop out at the end of the experiment.

Given the specified parameters and follow-up intervals, the data for an individual is firstly simulated by drawing random effects. Based on these effects, the bivariate longitudinal trajectories were simulated using the Metropolis-Hastings algorithm. Afterwards, random effects are used again to define the Gompertz parameter $\lambda^* = \exp(\beta + \alpha(\eta_0^{(1)} + b_{10}))$ for the first joint model and to calculate the survival function $S(t)$ for each follow-up time point. Using the cumulative distribution function $F(t) = 1 - S(t)$, the time point at death is derived by using the inversion method. When estimating the parameters, we use Gauss-Hermite quadrature to calculate the double integral. We choose 7 nodes for the corresponding Gauss-Hermite quadrature. The corresponding parameters are estimated using the *ucminf* function in package *ucminf*.⁴¹ The number of iterations is 100.

Performance measures: In this section, we will calculate the means of estimated parameters over the N_{sim} iterations. We will assess means, bias, relative bias (%), empirical standard error (SE), Monte Carlo SE of bias (MC.Bias) and Monte Carlo SE of empirical SE (MC.SE).⁴² Bias and relative bias are the main performance measures we are interested in. Monte Carlo SE is a statistic that quantifies the uncertainty of a simulation study with finite N_{sim} .

Table 1 shows the simulation results for the choices $N = 100, 400, 1000$, with follow-up intervals equals three. The number of iterations is 100. As we can observe in Table 1, in general, the mean of estimated values over the 100 iterations deviates little from the true values we set. The relative biases for most parameters are below 5%. It is worth noting the

Table 1. Simulation study for the first joint model. The follow-up interval is 3.

Parameter	Value	N = 100						N = 400						N = 1000					
		Mean	Bias	% bias	SE	MC.Bias	MC.SE	Mean	Bias	% bias	SE	MC.Bias	MC.SE	Mean	Bias	% bias	SE	MC.Bias	MC.SE
$\eta_0^{(1)}$	1.0	1.020	0.020	2.0	0.282	0.029	0.020	1.011	0.011	1.1	0.128	0.013	0.009	1.008	0.008	0.8	0.084	0.008	0.006
$\eta_1^{(1)}$	-0.2	-0.207	-0.007	3.5	0.046	0.005	0.003	-0.201	-0.001	0.7	0.020	0.002	0.001	-0.200	0.000	0.2	0.013	0.001	0.001
θ_{Y_1}	1.0	1.009	0.009	0.9	0.041	0.004	0.003	1.002	0.002	0.2	0.019	0.002	0.001	1.001	0.001	0.1	0.011	0.001	0.001
$\eta_0^{(2)}$	0.8	0.821	0.021	2.6	0.295	0.030	0.021	0.817	0.017	2.1	0.142	0.014	0.010	0.809	0.009	1.2	0.090	0.009	0.006
$\eta_1^{(2)}$	-0.2	-0.202	-0.002	1.2	0.044	0.004	0.003	-0.201	-0.001	0.4	0.020	0.002	0.001	-0.200	0.000	0.2	0.013	0.001	0.001
θ_{Y_2}	1.0	1.003	0.003	0.3	0.038	0.004	0.003	1.000	0.000	0.0	0.016	0.002	0.001	1.000	0.000	0.0	0.011	0.001	0.001
ϕ	0.8	0.798	-0.002	0.2	0.048	0.005	0.003	0.797	-0.003	0.3	0.022	0.002	0.002	0.798	-0.002	0.2	0.015	0.001	0.001
ξ	0.1	0.118	0.018	17.6	0.042	0.004	0.003	0.102	0.002	2.2	0.018	0.002	0.001	0.100	0.000	0.3	0.010	0.001	0.001
σ_{b1}	0.4	0.424	0.024	6.0	0.148	0.015	0.011	0.404	0.004	0.9	0.063	0.006	0.004	0.405	0.005	1.4	0.040	0.004	0.003
σ_{b2}	0.6	0.610	0.010	1.7	0.149	0.015	0.011	0.604	0.004	0.6	0.074	0.007	0.005	0.601	0.001	0.2	0.044	0.004	0.003
ρ	0.3	0.248	-0.052	17.5	0.393	0.040	0.028	0.308	0.008	2.6	0.173	0.017	0.012	0.315	0.015	5.0	0.113	0.011	0.008
β	-1.5	-1.500	0.000	0.0	0.603	0.062	0.044	-1.504	-0.004	0.3	0.254	0.025	0.018	-1.505	-0.005	0.4	0.145	0.015	0.010
α	-0.4	-0.453	-0.053	13.2	0.653	0.067	0.047	-0.410	-0.010	2.6	0.263	0.026	0.019	-0.399	0.001	0.1	0.146	0.015	0.010

MC.Bias: Monte Carlo standard error of bias; MC.SE: Monte Carlo standard error of empirical SE.

estimated values of the parameter ξ of the Gompertz hazard model, the correlation coefficient ρ of the random effects and the estimated value of the corresponding parameter α of the link function $\Delta(\alpha, \eta, \mathbf{b}, t)$. Although the estimates of these parameters have non-negligible biases when the sample size is small (the relative bias of these three parameters is 17.6%, 17.5% and 13.2%, respectively when the $N = 100$), the relative biases for all parameters are quite close to zero when the sample size is expanded to 1000. Overall, the mean of the estimated value will be closer to the set value when N is increased. As the sample size increases, there is a significant improvement in other relevant estimates such as bias, relative bias and SE.

Based on the bias values and the corresponding MC.Bias values in Table 1, we plot the Figure 1. From this figure, we can observe that for most of the parameters, 0 is included in the 95% confidence interval. From the left to the right of the figure, i.e. from $N = 100$ to $N = 1000$, the corresponding confidence intervals for the parameters' biases show a clear reduction. This means that the larger the sample size the closer the mean is to the true value, which is consistent with our previous discussion on Table 1.

This simulation also shows that

1. The estimation by marginal likelihood can reproduce the parameters that were used to generate the data.
2. The function *ucminf* combined with the Gauss-Hermite method gives accurate estimates. The settings of the arguments in this paper are reasonable, e.g. in the Gauss-Hermite method the node is set to 7 and in the simulation the number of iterations is set to 100. It is worth noting that while in this paper we have justified setting the number of nodes to 7 through simulation, in other cases, such as when there is considerable heterogeneity between subjects, we may need more nodes when using the Gauss-Hermite method.
3. The bias of the parameter estimates decreases significantly when the sample size increases. In particular, when the sample size is $N = 1000$, 0 is included in the 95% confidence interval of the estimated biases of all parameters. This result also justifies our choice of a sample size of 1000 in the 'Application section'.

5 Application

The ELSA is a rich resource of information on the dynamics of health, social, wellbeing and economic circumstances in the English population aged 50 and older. Established in 2002, the original sample was drawn from households that had

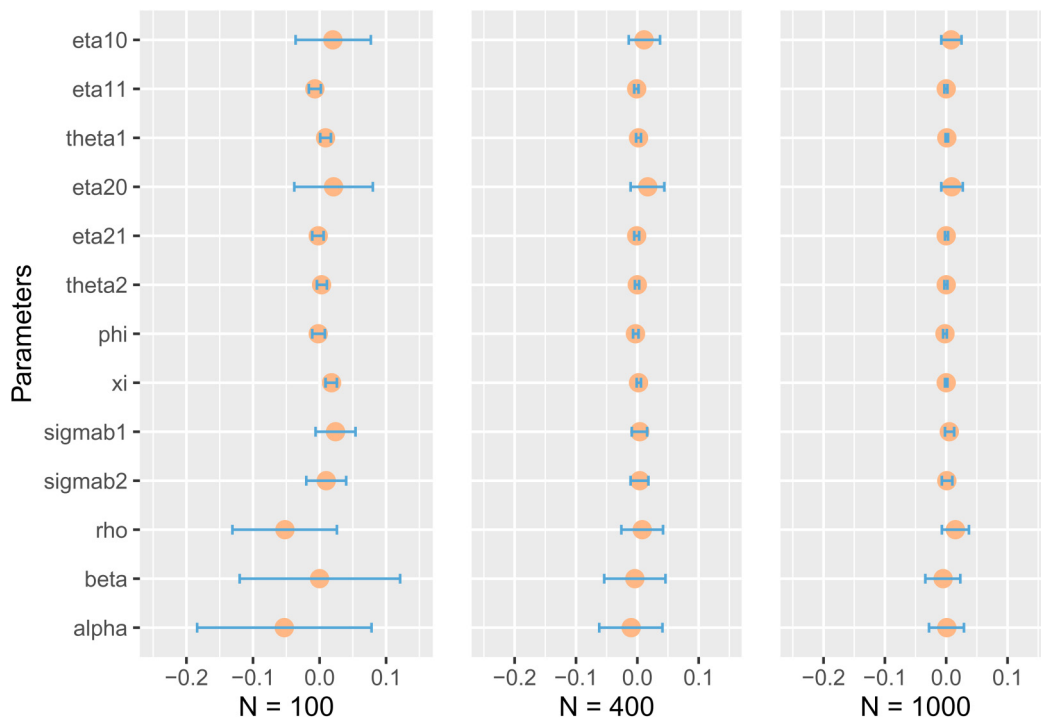


Figure 1. Bias and corresponding Monte Carlo 95% confidence interval. Circles represent biases, and geometry bars represent Monte Carlo 95% confidence interval.

previously responded to the Health Survey for England between 1998 and 2001. The same group of respondents have been interviewed at two-yearly interviews, known as ‘waves’, to measure changes in their health, economic and social circumstances. Younger age groups are replaced or refreshed to retain the panel. Data from the ELSA can be obtained via the Economic and Social Data Service (www.esds.ac.uk). The information collected provides data on household and individual demographics, health, social care, work and pensions, income and assets, housing, cognitive function, social participation, effort and reward, expectations, walking speed and weight for individuals.

In this paper, we focus on a test with two responses in the ELSA data.⁴³ This test asks individuals to remember 10 words in the same interview. The first response in the test, representing the individual’s short-term memory ability to some extent, is the number of words the individual recalls immediately (immediate recall). An individual’s long-term memory ability is indicated by recording the number of words he/she can remember after five minutes (delayed recall).

The number of immediate-recall words is represented by Y_1 and that for the delayed-recall words is Y_2 , in which $Y_j \in \{0, 1, \dots, 10\}, j = 1, 2$. The corresponding time (age) is given in integers for the reason of confidentiality.⁴³

In the ELSA data, there are 11,932 individuals interviewed. In this section, we analyze the ELSA data which were also used in Van den Hout and Muniz-Terrera’s paper.⁴ Van den Hout and Muniz-Terrera processed the data via the following four aspects: remove individuals who were interviewed only once with missing data on the number of words; remove individuals without information on the year of birth; remove individuals who were younger than 50 years old at baseline wave 1 and remove individuals with censored age at baseline.

In this paper, the main purpose of this paper is to illustrate the methodology rather than a study of ELSA data. Therefore, we use a subset of the data above.⁷ The subset has $N = 1000$ individuals, randomly sampled from the full data conditional on two restrictions:

1. Individuals are younger than 90 years at baseline and do not have censored age during follow-up.
2. Each individual has at least two records. These two records can either include two observations or an observation and a time of death.

The age in this dataset is rescaled by subtracting 49. The number of deaths is 195, where the attrition rate is close to the full data. The ratio of women to men in the data is 540: 460, which is also close to that ratio for the full data. The average of first recorded immediate-recall words for all individuals is 5.60, and that for the delayed recall is 4.19. Other details for this subset can be found in Van Den Hout.³⁹

Figure 2 shows the frequency distribution of recalled words at baseline interview. Compared with the distribution of immediate recall words, the distribution of delayed-recall words is concentrated on the left. It can also reflect that individual’s short-term memory is better than long-term memory. In Figure 3, we select 30 individuals and plot the trajectories of their responses. We can see from the picture that an individual’s ability to remember words (cognitive ability) is fluctuating and overall the number of words that people remember decreases over time.

The joint model in this subsection uses the extension of bivariate binomial distribution mentioned in the the ‘Longitudinal model section; the Gompertz hazard model and the Weibull hazard model mentioned in the the ‘Hazard models section. The longitudinal model and the survival model are joined together given the normally distributed bivariate random effects. The joint model mentioned in this section is divided into two main parts: the first part is a bivariate

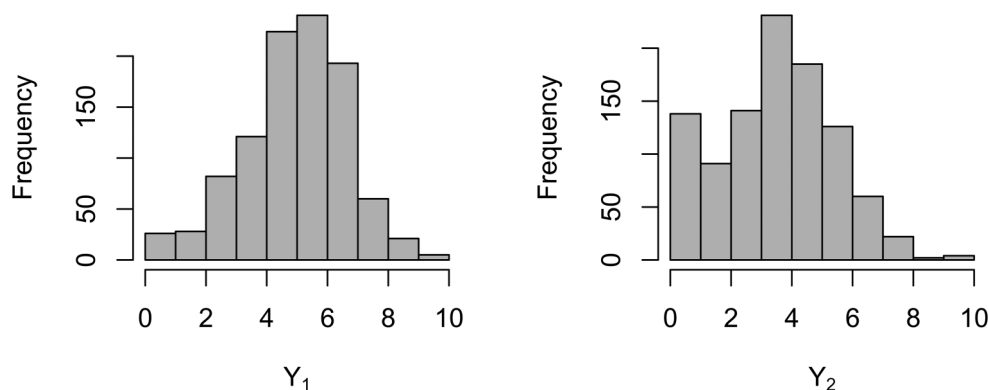


Figure 2. Frequency distribution of recalled words at baseline interview. Y_1 represents immediate-recall words; Y_2 represents delayed-recall words.

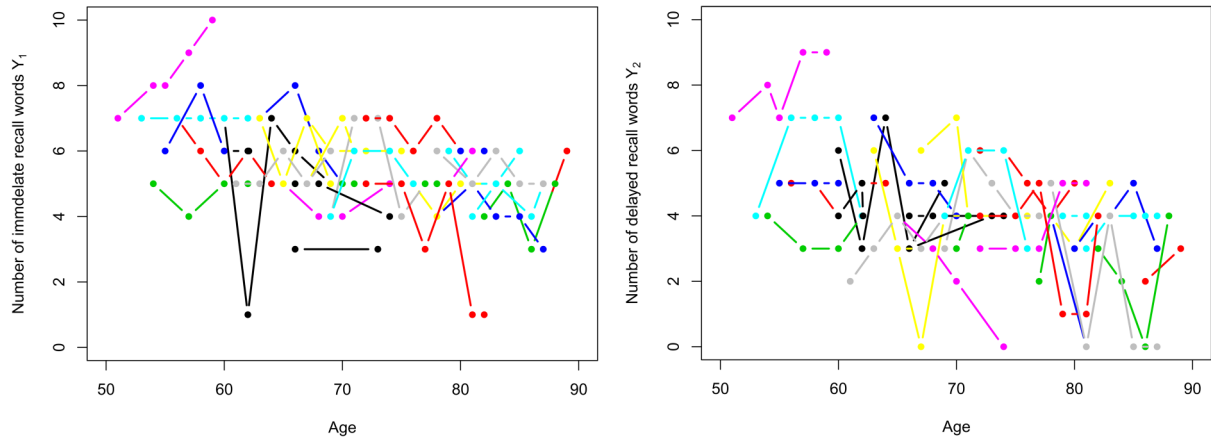


Figure 3. Recalled words trajectories. Left-hand side: immediate-recall words Y_1 trajectories. Right-hand side: delayed-recall words Y_2 trajectories. Individuals are represented in the same colour in both plots.

Table 2. ELSA: Parameter estimates for the bivariate extension of the joint model in paper.⁴ The values in parentheses are the standard errors of the corresponding parameters.

$\eta_0^{(1)}$	-0.857 (0.053)	θ_{Y_1}	1.122 (0.009)	σ_{b1}	1.116 (0.073)	β	-6.800 (0.497)
$\eta_0^{(2)}$	-1.969 (0.094)	θ_{Y_2}	1.167 (0.011)	σ_{b2}	0.055 (0.005)	ξ	0.097 (0.012)
$\eta_1^{(1)}$	-0.014 (0.002)	ϕ	1.432 (0.016)	ρ	-0.682 (0.190)		
$\eta_1^{(2)}$	-0.038 (0.004)			α_1	-0.400 (0.168)	α_2	-0.193 (0.086)

extension model based on the univariate model proposed in Van den Hout and Muniz-Terrera's paper⁴; the second part contains joint models constructed based on the link functions we discussed in the 'Link between the models section.

5.1 Bivariate extension of joint model based on previous paper⁴

Van den Hout and Muniz-Terrera constructed a univariate shared random-effect joint model with binomial distribution and beta-binomial distribution.⁴ Here, we would like to use a similar method, i.e. a shared random-effect method, to construct bivariate joint models. Although we use different distributions and have different numbers of responses, because the same shared random-effect method is applied, we think it is still worthwhile to construct a bivariate model with a similar structure based on this univariate model in Van Den Hout and Muniz-Terrera.⁴ We use the same structure with random intercept and random slope as in Van Den Hout and Muniz-Terrera⁴ for one of the responses, and the other response is modelled with fixed effects.

For the univariate joint model in Van Den Hout and Muniz-Terrera⁴, the estimation of α s are negative. If the estimation of α s in the bivariate joint model is similar to the α s for the univariate joint model, i.e. both α s are negative, we can infer that our method is consistent with the analysis based on the univariate model. Since the response variable of the univariate joint model is delayed recall Y_2 , we add the random effects to the linear predictor associated with Y_2 . We assume p_{Y_2} equals to (6), where $j = 2$. For immediate recall Y_1 , we have the logistic regression model p_{Y_1} :

$$p_{Y_1} = \frac{\exp(\eta_0^{(1)} + \eta_1^{(1)}t)}{1 + \exp(\eta_0^{(1)} + \eta_1^{(1)}t)} \quad (30)$$

The corresponding link function is

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \mathbf{b}, t) = \alpha_1(\eta_0^{(2)} + b_{20}) + \alpha_2(\eta_1^{(2)} + b_{21})t \quad (31)$$

The hazard model also contains the same link function for delayed recall. The hazard model equals to (25). Parameter λ^* and ξ^* in the hazard model equal to (26) for $j = 2$. Estimated result are provided in Table 2.

The α s' estimates in Table 2 meet our expectation. The estimates $\hat{\alpha}_1 = -0.400$ is negative and significantly different from zero, which means that the risk of death is relatively lower when the individual has a better cognitive function at the baseline age. Meanwhile, the parameter $\hat{\alpha}_2 = -0.193$ is also negative, implying that the risk of death increases with time given that b_{21} is equal to 0 on average.

The probability of remembering a word at baseline age ($t = 0$), conditional on random effects being equal to its mean 0, could be calculated via parameter $\eta_0^{(j)}, j = 1, 2$. For the immediate-recall number, the corresponding we have $\hat{\eta}_0^{(1)} = -0.857$ and the corresponding probability equals to 0.298; for the number of delayed recall, the corresponding probability decreases to 0.122 with $\hat{\eta}_0^{(2)} = -1.969$. The probability of remembering a word immediately at baseline age is higher than that of recalling later. It means that the expected immediate-recall number is larger than the expected number of the delayed recall. Both $\hat{\eta}_1^{(1)}$ and $\hat{\eta}_1^{(2)}$ are smaller than 0, which shows that individuals' memory or cognitive function declines with age. Moreover, since $\hat{\eta}_1^{(1)} > \hat{\eta}_1^{(2)}$, we can infer that for each individual the transient memory decays more slowly than long-term memory with age.

5.2 Bivariate extension joint models with different link functions Δ

We will now construct the joint model based on the link function mentioned in the 'Link between the models section. Our joint model consists of the bivariate extension binomial distribution as the longitudinal model and the Gompertz hazard or the Weibull hazard as the survival model. Gender is added to the model as a covariate (0 for women, 1 for men). The corresponding logistic regression model and link function is given by

$$p_{Y_j} = \frac{\exp\left((\eta_0^{(j)} + b_{j0}) + (\eta_1^{(j)} + b_{j1})t + \gamma_L^{(j)}sex\right)}{1 + \exp\left((\eta_0^{(j)} + b_{j0}) + (\eta_1^{(j)} + b_{j1})t + \gamma_L^{(j)}sex\right)} \tag{32}$$

where $j = 1$ or 2 .

We also add gender to the link function $\Delta(\alpha, \eta, \mathbf{b}, \gamma_S, t)$. Taking equation (16) as an example, in this section we write equation (16) as:

$$\Delta(\alpha, \eta, \mathbf{b}, \gamma_S, t) = \sum_{j=1}^2 \alpha_j \left(\eta_0^{(j)} + b_{j0} \right) + \gamma_S sex \tag{33}$$

Because we want to distinguish the effect of two responses on the risk of death, in each joint model we set only one random effect for one response. It means that we only discuss the case where both responses have random intercepts or both have random slopes. Moreover, in the joint model, we parameterize the correlation between the two responses multiple times using four parameters: ϕ in bivariate extension binomial model (4), correlation ρ between corresponding random effects for two responses and link function parameters α_1 and α_2 . We will discuss the case in which α_1, α_2 and ρ are equal to 0 respectively.

We use *INT* to denote the joint model containing random intercept in the link function, i.e. link function equals to (15) or (16). In this case for the longitudinal model we restrict $b_{11} = b_{21} = 0$ (equation (9)). We use *SLO* to represent models with random slope in the link function. In this situation, the corresponding link function is equal to either (18) or (17), and the corresponding longitudinal are restricted to equation (10). The letters *G* and *W* denote models containing the Gompertz hazard model and the Weibull hazard model, respectively.

The Akaike information criterion (AIC) values and the corresponding parameters of link function Δ are presented in Table 3. Models *INT.W1* - 4 and Model *INT.G1* - 4 in Table 3 represent joint models with random intercept. For Model *SLO.G1* - 4, we construct the joint model with random slope.

Table 3. ELSA: AIC and estimated α s for shared random-effects joint models. $-2LL$ represents the $-2\log$ likelihood.

MODEL	p_{Y_j}	Link function	Hazard distribution	$-2LL$	AIC	α	α_1	α_2	ρ
<i>INT.W1</i>		(16)	Weibull	100985.1	101019.1		-0.480	-0.031	0.316
<i>INT.W2</i>		(15) $j=1$	Weibull	100985.2	101017.2	-0.508			0.318
<i>INT.W3</i>		(15) $j=2$	Weibull	100989.2	101021.2	-0.154			0.327
<i>INT.W4</i>	(9)	(16)	Weibull	101001.2	101033.2		-0.507	-0.075	
<i>INT.G1</i>		(16)	Gompertz	100965.5	100999.5		-0.609	-0.007	0.315
<i>INT.G2</i>		(15) $j=1$	Gompertz	100965.5	100997.5	-0.616			0.316
<i>INT.G3</i>		(15) $j=2$	Gompertz	100971.2	101003.2	-0.168			0.329
<i>INT.G4</i>		(16)	Gompertz	100981.5	101013.5		-0.632	-0.065	
<i>SLO.G1</i>	(10)	(17)	Gompertz	103373.2	103407.2		0.062	-0.015	0.474
<i>SLO.G2</i>		(18) $j=1$	Gompertz	101170.6	101202.6	-0.617			0.436
<i>SLO.G3</i>		(18) $j=2$	Gompertz	101179.1	101211.1	-0.117			0.431
<i>SLO.G4</i>		(17)	Gompertz	101197.5	101229.5		-0.583	-0.069	

ELSA: English Longitudinal Study of Ageing; AIC: Akaike information criterion.

We expect the risk of death to be relatively lower when individuals have a good cognitive function at baseline age or the downward trajectory over time of cognitive function is slow. For all models except Model *SLO.G1* in Table 3, the α s are estimated as smaller than 0. For Models *INT.W1* - 4 and *INT.G1* - 4, the negative α s mean that the risk of death is relatively lower when the individual has a better cognitive function at the baseline age. For Models *SLO.G2*, *SLO.G3* and *SLO.G4*, the risk of death is lower when the cognitive function declines relatively slower for an individual.

When the link function has random slopes for both responses (17), the model's estimate of α_1 is not in line with our expectations with respect to the process. The estimated parameter $\hat{\alpha}_1 = 0.062$ in Model *SLO.G1* is positive. This positive α_1 , representing the individual risk of death becomes smaller over time with the corresponding random slope of immediate recall. This result is the opposite of what we expected: instead of the risk of death increasing over time, it decreases over time.

In general, the AIC value for the model with random intercept (e.g. Model *INT.W2*) is smaller than that for the corresponding model with random slope (e.g. Model *SLO.W2*). It means that there was little heterogeneity in slope between subjects, so the increase in model likelihood was negligible compared to the increase in complexity.

• **Random intercept for both Y_1 and Y_2**

The joint Model *INT.G2*, using the Gompertz survival model with link function (15) for $j = 1$, fits the best. We present the parameter estimation in Table 4.

In Table 4, we have $\hat{\eta}_0^{(1)} = -0.580$ is larger than $\hat{\eta}_0^{(2)} = -2.014$, we can calculate the corresponding probability of remembering a word at baseline age via these two parameters. The probability of recalling a word immediately at baseline age given $b = 0$ is 0.359, in which is higher than that for recalling a word later (0.118). The high probability of remembering words implies better expected cognitive function. For an individual, the expected immediate-recall number is larger than that of the delayed recall. Meanwhile, the estimation of $\hat{\alpha} = -0.616$ is negative, which means that the risk of death is relatively lower when the individual has a better cognitive function. Based on the estimated SE of α , α is significantly different from zero. This implies that there is a link between the longitudinal process and the survival process. Moreover, both $\hat{\eta}_1^{(1)}$ and $\hat{\eta}_1^{(2)}$ are smaller than 0, which means that the probability of recalling a word decreases with age. We also notice that random effects b_1 and b_2 have a positive correlation with $\hat{\rho} = 0.370$. It means that if the individual has great short-term memory (immediately recall ability) at the beginning of the research, the corresponding long-term memory (delayed recall ability) will be slightly positively affected. Parameters $\hat{\theta}_{Y_1}$ and $\hat{\theta}_{Y_2}$ are both larger than 1, which means that the estimated distribution is more peaked than the standard bivariate binomial distribution. The three parameters related to gender, $\hat{\gamma}_L^{(1)}$ and $\hat{\gamma}_L^{(2)}$ in the longitudinal model are both negative. This result means that, with other parameters being equal, men are expected to have lower cognitive function than women. Parameter $\hat{\gamma}_S = 0.370$ implies a higher risk of death for men than for women under the same circumstances.

• **Additional remarks**

Compare with the probability of remembering a word ($\eta_0^{(1)}$ and $\eta_0^{(2)}$) at baseline level, we expect that change in cognitive ability over time will have a greater impact on the risk of death.

For the Gompertz hazard model, we rewrite the hazard model based on (25):

$$\begin{aligned}
 h(t) &= \exp(\beta + \eta_0^{(j)} + \alpha(\eta_1^{(j)} + b_{j1})t + \xi t) \\
 &= \exp\left\{\beta + \eta_0^{(j)} + (\alpha(\eta_1^{(j)} + b_{j1}) + \xi)t\right\}
 \end{aligned}
 \tag{34}$$

In this case, the parameter ξ^* for the Gompertz model is:

$$\xi^* = \alpha(\eta_1^{(j)} + b_{j1}) + \xi.
 \tag{35}$$

In Table 3, we constructed four models which use random slope as link function (Model *SLO.G1* - 4). The joint Model *SLO.G2*, using the Gompertz survival model with link function (18) for $j = 1$, has the best fitting effect among the four models.

Table 4. ELSA: Parameter estimates for Model *INT.G2* with (15) $j = 1$. The values in parentheses are the standard errors of the corresponding parameters.

$\eta_0^{(1)}$	-0.580 (0.080)	θ_{Y_1}	1.206 (0.035)	σ_{b1}	0.447 (0.101)	β	-7.043 (0.350)
$\eta_0^{(2)}$	-2.014 (0.108)	θ_{Y_2}	1.123 (0.028)	σ_{b2}	0.649 (0.094)	ξ	0.122 (0.009)
$\eta_1^{(1)}$	-0.024 (0.002)	ϕ	1.425 (0.061)	ρ	0.316 (0.082)		
$\eta_1^{(2)}$	-0.029 (0.003)					α	-0.616 (0.227)
$\gamma_L^{(1)}$	-0.063 (0.042)	$\gamma_L^{(2)}$	-0.130 (0.051)	γ_S	0.370 (0.148)		

In Model *SLO.G2*, the estimation of $\hat{\eta}_0^{(1)} = -0.631$ is higher than $\hat{\eta}_0^{(2)} = -1.956$, representing the higher probability of remembering a word. The expected immediate-recall number for an individual is greater than that of the delayed recall. The estimation of $\hat{\alpha} = -0.617$ is negative, representing the relatively lower risk of death when the individual has a better cognitive function. Estimated $\hat{\eta}_1^{(1)}$ and $\hat{\eta}_1^{(2)}$ are smaller than 0, which means that the probability of recalling a word decreases with age. Similarly with Table 4, the correlation between random effects b_1 and b_2 is positive ($\hat{\rho} = 0.436$). This time, it shows that if an individual's instant memory declines rapidly with age, his/her long-term memory may also decline fast.

In Table 3, we have $\hat{\alpha} > 0$ when we include random effects for both Y_1 and Y_2 in the survival model. These estimates represent that higher cognitive function leads to a high risk of death. The estimated α s are not affected by the type of random effects (intercept/slope). Although the AIC value for the joint model *SLO.G1* is not the smallest, the values of $\hat{\alpha}_1$ is still worth to be analyzed. This estimation may be caused by the correlation between Y_1 and Y_2 . In the model in this paper, we captured the relationship between the two responses several times via parameters ϕ , ρ and α s. Overfitting in this model may lead to unexpected results.

5.3 Model fit

In order to investigate model fit, in this section we will first calculate the corresponding individual-specific random effects and then plot the fitted distribution based on the mean of random effects. After that we will discuss the accuracy of model predictions by comparing the predicted survival curves with the observed survival curves.

Denote the responses for individual i by \tilde{y}_i . Let \tilde{t}_{in_i} denote the age corresponding to the last observed data, and \tilde{t}_{i1} denote the age corresponding to the baseline age. When there is only one observation, we set $\tilde{t}_{in_i} = \tilde{t}_{i1}$. In order to predict an individual trajectory given observed data, we need the corresponding values of random effects. We use Maximum a posteriori (MAP) estimation to derive the most likely value of the random effects:

$$p(\mathbf{b}_i | \tilde{t}_{i1}, \tilde{t}_{in_i}, \delta_i = 0, \tilde{y}_i; \omega = \hat{\omega}) \propto p(\tilde{t}_{in_i} | \delta_i = 0, \mathbf{b}_i; \omega = \hat{\omega}) p(\tilde{y}_i | \mathbf{b}_i; \omega = \hat{\omega}) p(\mathbf{b}_i | \omega = \hat{\omega}) \quad (36)$$

One disadvantage of shared random-effects model is that we define the distribution of random effect before fitting models. We can verify whether the distribution of random variables is the same as the preset distribution, which is bivariate normal distribution in this paper, based on the estimated random effects. The estimated random effects for model in the 'Bivariate extension of joint model based on previous paper[4] section follows bivariate normal distribution $\mathbf{b}_i \sim N\left(\begin{pmatrix} -0.017 \\ 0.002 \end{pmatrix}, \begin{pmatrix} 0.306 & -0.345 \\ -0.345 & 0.001 \end{pmatrix}\right)$. The expectations of random effects are close to what we set, and the estimated variances are smaller than the result in Table 2. Based on Nan Laird's book,⁴⁴ this result is in line with our expectation. Given estimated random effects, both survival and longitudinal responses can be predicted given estimated random effects and an assumed age.

We draw the three-dimensional plot Figure 4 of the bivariate binomial distribution based on the result in Table 2 with random effect \hat{b}_i s equal to the mean of estimated \mathbf{b}_i in (36). The left-hand side density plot is conditional on age 50, and the right-hand side density is conditional on age 80. It is clear that as age increases, the individual's number of recalled words are likely to decrease. At the age of 50, the number of immediate-recall words is concentrated in 4–10. When the age reaches 80, the number of immediate-recall words is reduced to 2–8. This change is more obvious in the number of delayed-recall words. At the age of 50, the number of delayed-recall words is concentrated in 3–8. However, the number of delayed-recall words for most individuals drops sharply to 0–5 when individuals' age reaches 80. Numerically, we can confirm this conclusion via the corresponding expectations: $E(Y_1, Y_2 | t = 50) = (6.8, 5.6)$; $E(Y_1, Y_2 | t = 80) = (4.6, 3.0)$.

Given the observed age and corresponding responses, we expect that our joint model can predict the risk of death. To prove this, we plot predicted survival curves based on the calculated random effects and compared these curves to the observed survival curve (i.e. the K-M curve). Figure 5 shows the comparison of Kaplan–Meier survival curves and predicted model survival curves based on Model *INT.G2* with (15) $j = 1$. The blue line shows the mean of predicted survival curves, and the red line shows the median of predicted survival curves. Only the baseline test score is used in Figure 5. The comparison plot illustrates that our joint model has a relatively better predictive ability for individual risk of death.

6 Conclusion

Joint models are constructed to analyze longitudinal and time-to-event data in ageing research. Since the tests used in the application are based on discrete values, and we are interested in two tests, the bivariate extension of the binomial distribution with mixed-effects regression is specified to model the longitudinal data. The bivariate extension of the binomial

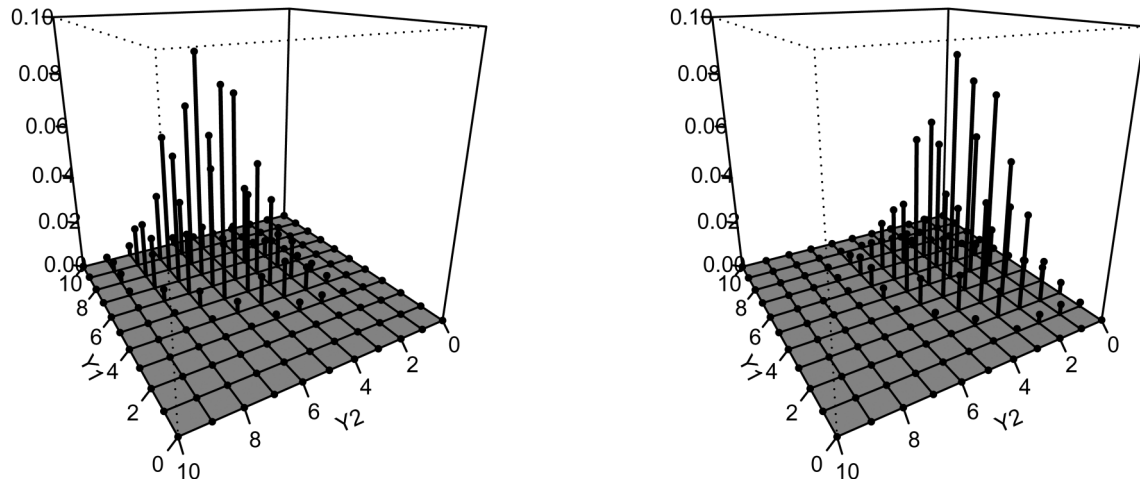


Figure 4. Fitted bivariate binomial distribution for immediate recall (Y_1) and delayed recall (Y_2). Left hand side: conditional on age 50; Right hand side: conditional on age 80.

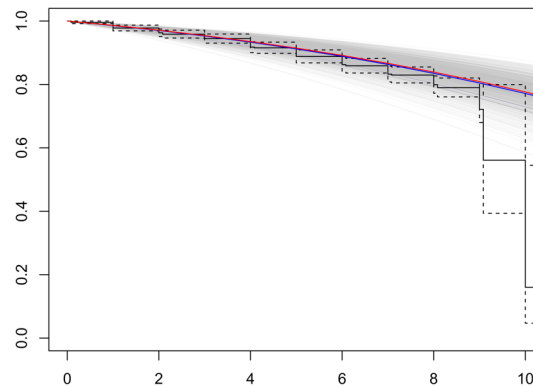


Figure 5. Comparison of K-M survival curves: predicted survival curves (grey lines), the mean of those survival curves (blue line) and the median of survival curves (red line).

distribution allows modelling for correlation between the responses. The Gompertz hazard model and the Weibull hazard model are explored for the time-to-event data. The two models are joined together via the link function Δ , which contains the random effect b . The random effect is used to capture the dependence across time and the dependence between responses.

The joint model is used to analyze the ELSA data. Model comparison is undertaken by applying the Akaike information criterion. The AIC of the joint model with random intercept link function is smaller than the joint model using random slope link function, based on the same expression. When the link function equals to random slope for both responses, we got unexpected results: the slowly decline of the cognitive function over time is correlated with the high risk of death or dementia. The cause of this phenomenon needs further investigation.

We used the shared random-effects model to construct joint models in this paper. This model offers a flexible way to measure the correlation between the responses and the risk of event, i.e. the random effect follows a distribution instead of being equated to a fixed value. Instead of using random effects to capture correlation between responses, the joint model in this paper uses a separate parameter to model the correlation between responses. This allows us to more clearly identify the relationship between two responses.

Our modelling framework has some disadvantages. Firstly, it is computationally intensive since we need to calculate the double integral. Secondly, the distribution of random effects is set in advance. For example, we use the bivariate normal distribution in this paper. The multivariate normal distribution is the standard modelling choice for random effects in longitudinal submodels.³ Pantazis and Touloumi⁴⁵ investigated the misspecification of the random effects distributions. They conclude that the fixed effects parameter estimates are fairly robust, except for the parameter estimates in the time-to-event

sub-model. However, the SEs may be underestimated for severely skewed distribution. If the actual distribution is different from the normal distribution, it may lead to biased estimates.

There are still some aspects that can be used as starting points for further research. Firstly, we use linear predictors and logistic regression to model the probability in the bivariate binomial distribution. In further research, the linear effect of age in the logistic regression can be modelled more flexibly by using B-splines or other semi-parametric methods. Secondly, there are restrictions imposed by the shared random-effects model, which we have discussed earlier in this paper. Thirdly, in this paper, we construct a bivariate joint model whereas in longitudinal data there may be more than two responses. It is possible to construct the corresponding multivariate extended binomial distribution based on Altham and Hankin,³⁰ and use this distribution in our joint model. However, this will imply a computational challenge with respect to integrating out the random effects. An alternative might be to use a Bayesian approach; for a recent development in this area see the *JMbayes* package.²⁵ Lastly, as we mentioned in the ‘Models section, our framework is flexible in its choice of parametric hazard models, i.e. it is easy to replace the Gompertz hazard model and the Weibull hazard model by other parametric hazard models. Joint models using other parametric hazard models are worth being investigated.

To conclude, in this paper we construct a bivariate joint model to investigate the change in cognitive function. This bivariate joint model is general and can be used in a wide range of bivariate discrete-valued tests in ageing research. We discussed the shared part which is contained in both the longitudinal model and survival model. We hope our discussion will provide ideas for constructing other bivariate joint models in ageing research.

Acknowledgements

The authors really appreciate researchers in the University College London, NatCen Social Research, and the Institute for Fiscal Studies, who collect the ELSA data and make it available to PhD students and researchers. The ELSA was developed by a team of researchers based at the University College London, NatCen Social Research, and the Institute for Fiscal Studies. The data were collected by NatCen Social Research. The funding is currently provided by the National Institute of Aging (R01AG017644), and a consortium of UK government departments coordinated by the National Institute for Health Research.

The authors would like to thank the reviewers for their comments on our manuscript. We have learned a great deal from their feedback, which has enabled us to improve the manuscript considerably and allowed us to take note of relevant issues in our future research work.

Availability of data and materials

The authors do **not** have permission to share the data, but people can apply for access through the ELSA website. We are happy to provide the code, especially the code for the ‘Simulation section. If you would like to get the code used in the paper please contact the corresponding author.


Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Shengning Pan  <https://orcid.org/0000-0001-8895-9993>

References

1. Diggle PJ, Heagerty P, Liang KY et al. *Analysis of Longitudinal Data*. Oxford: Oxford University Press, 2002.
2. Banks J, Breeze E, Lessof C et al. Retirement, health and relationships of the older population in England: The 2004 English longitudinal study of ageing (wave 2), 2006.
3. Hickey GL, Philipson P, Jorgensen A et al. Joint modelling of time-to-event and multivariate longitudinal outcomes: Recent developments and issues. *BMC Med Res Methodol* 2016; **16**: 1–15.
4. Van Den Hout A and Muniz-Terrera G. Joint models for discrete longitudinal outcomes in aging research. *J R Stat Soc: Ser C (Applied Statistics)* 2016; **65**: 167–186.
5. Laird NM and Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**: 963–974.
6. Cox DR. Regression models and life-tables. *J R Stat Soc: Ser B (Methodological)* 1972; **34**: 187–202.
7. Van Den Hout A and Muniz-Terrera G. Hidden three-state survival model for bivariate longitudinal count data. *Lifetime Data Anal* 2019; **25**: 529–545.

8. Ibrahim JG, Chu H and Chen LM. Basic concepts and methods for joint models of longitudinal and survival data. *J Clin Oncol* 2010; **28**: 2796.
9. Dupuy Jf and Mesbah M. Joint modeling of event time and nonignorable missing longitudinal data. *Lifetime Data Anal* 2002; **8**: 99–115.
10. Hogan JW and Laird NM. Increasing efficiency from censored survival data by using random effects to model longitudinal covariates. *Stat Methods Med Res* 1998; **7**: 28–48.
11. Henderson R, Diggle P and Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; **1**: 465–480.
12. Mauff K, Steyerberg E, Kardys I et al. Joint models with multiple longitudinal outcomes and a time-to-event outcome: A corrected two-stage approach. *Stat Comput* 2020; **30**: 999–1014.
13. Touloumi G, Pantazis N, Babiker AG et al. Differences in hiv rna levels before the initiation of antiretroviral therapy among 1864 individuals with known hiv-1 seroconversion dates. *Aids* 2004; **18**: 1697–1705.
14. Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Boca Raton: Chapman and Hall/CRC, 2012.
15. Philipson P, Diggle P, Sousa I et al. *joineR: Joint modelling of repeated measurements and time-to-event data*, 2012. <https://github.com/graemeleehickey/joineR/>. R package version 1.2.6.
16. Crowther MJ, Abrams KR and Lambert PC. Joint modeling of longitudinal and survival data. *Stata J* 2013; **13**: 165–184.
17. Ibrahim JG, Chen MH and Sinha D. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Stat Sin* 2004; **14**: 863–883.
18. van Boven N, Battes LC, Akkerhuis KM et al. Toward personalized risk assessment in patients with chronic heart failure: Detailed temporal patterns of nt-probnp, troponin t, and crp in the bio-shift study. *Am Heart J* 2018; **196**: 36–48.
19. Hatfield LA, Boye ME and Carlin BP. Joint modeling of multiple longitudinal patient-reported outcomes and survival. *J Biopharm Stat* 2011; **21**: 971–991.
20. Huang W, Zeger SL, Anthony JC et al. Latent variable model for joint analysis of multiple repeated measures and bivariate event times. *J Am Stat Assoc* 2001; **96**: 906–914.
21. Rizopoulos D and Ghosh P. A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Stat Med* 2011; **30**: 1366–1380.
22. Wang C, Douglas J and Anderson S. Item response models for joint analysis of quality of life and survival. *Stat Med* 2002; **21**: 129–142.
23. Rue M, Andrinopoulou ER, Alvares D et al. Bayesian joint modeling of bivariate longitudinal and competing risks data: An application to study patient-ventilator asynchronies in critical care patients. *Biom J* 2017; **59**: 1184–1203.
24. Crowther MJ. merlin—a unified modeling framework for data analysis and methods development in stata. *Stata J* 2020; **20**: 763–784.
25. Rizopoulos D, Papageorgiou G and Miranda Afonso P. *JMbayes2: Extended Joint Models for Longitudinal and Time-to-Event Data*, 2022. <https://drizopoulos.github.io/JMbayes2/>, <https://github.com/drizopoulos/JMbayes2>.
26. Proust-Lima C, Joly P, Dartigues JF et al. Joint modelling of multivariate longitudinal outcomes and a time-to-event: a nonlinear latent class approach. *Comput Stat Data Anal* 2009; **53**: 1142–1154.
27. Lin H, McCulloch CE, Turnbull BW et al. A latent class mixed model for analysing biomarker trajectories with irregularly scheduled observations. *Stat Med* 2000; **19**: 1303–1318.
28. Rouanet A, Joly P, Dartigues JF et al. Joint latent class model for longitudinal data and interval-censored semi-competing events: Application to dementia. *Biometrics* 2016; **72**: 1123–1135.
29. Proust-Lima C, Dartigues JF and Jacqmin-Gadda H. Joint modeling of repeated multivariate cognitive measures and competing risks of dementia and death: A latent process and latent class approach. *Stat Med* 2016; **35**: 382–398.
30. Altham PM, Hankin RK et al. Multivariate generalizations of the multiplicative binomial distribution: Introducing the mm package. *J Stat Softw* 2012; **46**: 1–23.
31. Diggle P, Henderson R and Philipson P. Random-effects models for joint analysis of repeated-measurement and time-to-event outcomes. In: Fitzmaurice G, Davidian M, Verbeke G, et al. (eds) *Longitudinal data analysis*. 1st ed. Chapman and Hall/CRC, 2008. DOI: 10.1201/9781420011579.
32. McCullagh P and Nelder J. *Generalized linear models*, 1989.
33. Altham PM. Two generalizations of the binomial distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 1978; **27**: 162–167.
34. Liu X. *Survival analysis: Models and applications*. Now York: John Wiley & Sons, 2012.
35. Pollard JH and Valkovics EJ. The gompertz distribution and its applications. *Genus* 1992; **48**: 15–28.
36. Kleinbaum DG and Klein M. *Survival Analysis*. 3. New York: Springer, 2010.
37. Wienke A. *Frailty Models in Survival Analysis*. Boca Raton: Chapman and Hall/CRC, 2010.
38. Crowther MJ, Abrams KR and Lambert PC. Flexible parametric joint modelling of longitudinal and survival data. *Stat Med* 2012; **31**: 4456–4471.
39. Van Den Hout A. *Multi-state survival models for interval-censored data*. Boca Raton: CRC Press, 2017.
40. Team RC et al. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2013.

41. Nielsen HB and Mortensen SB. *ucminf: General-Purpose Unconstrained Non-Linear Optimization*, 2016. <https://CRAN.R-project.org/package=ucminf>. R package version 1.1-4.
42. Morris TP, White IR and Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med* 2019; **38**: 2074–2102.
43. Taylor R, Conway L, Calderwood L et al. Health, wealth and lifestyles of the older population in England: The 2002 English longitudinal study of ageing technical report. London: Institute of Fiscal Studies 2007. https://sp.ukdataservice.ac.uk/doc/5050/mrdoc/pdf/5050_Wave_1_Technical_Report.pdf.
44. Laird N. Analysis of longitudinal and cluster-correlated data. *NSF-CBMS Regional Conference Series in Probability and Statistics* 2004; **8**: i–155. <http://www.jstor.org/stable/4153193>.
45. Pantazis N and Touloumi G. Robustness of a parametric model for informatively censored bivariate longitudinal data under misspecification of its distributional assumptions: a simulation study. *Stat Med* 2007; **26**: 5473–5485.