

A mixture model for the joint analysis of latent developmental trajectories and survival

Rinke H. Klein Entink,^a Jean-Paul Fox,^{b,*†} and Ardo van den Hout^c

A general joint modeling framework is proposed that includes a parametric stratified survival component for continuous time survival data, and a mixture multilevel item response component to model latent developmental trajectories given mixed discrete response data. The joint model is illustrated in a real data setting, where the utility of longitudinally measured cognitive function as a predictor for survival is investigated in a group of elderly persons. The object is partly to determine whether cognitive impairment is accompanied by a higher mortality rate. Time-dependent cognitive function is measured using the generalized partial credit model given occasion-specific mini-mental state examination response data. A parametric survival model is applied for the survival information, and cognitive function as a continuous latent variable is included as a time-dependent explanatory variable along with other explanatory information. A mixture model is defined, which incorporates the latent developmental trajectory and the survival component. The mixture model captures the heterogeneity in the developmental trajectories that could not be fully explained by the multilevel item response model and other explanatory variables. A Bayesian modeling approach is pursued, where a Markov chain Monte Carlo algorithm is developed for simultaneous estimation of the joint model parameters. Practical issues as model building and assessment are addressed using the DIC and various posterior predictive tests. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: joint longitudinal and survival modeling; MCMC; mixture model; MMSE; developmental trajectory modeling; Bayesian item response modeling

1. Introduction

In medical and epidemiological research, questionnaire data are often used to measure a patient's status of health or functioning (cognitive or physical). The multiple response observations that are recorded using a standardized test are used to measure a continuous latent variable. This continuous latent variable might represent disability in daily living [1], health-related quality of life [2, 3] or cognitive impairment [4]. The variation in the latent variable might not be fully explained by the usual measurement model when dealing with a highly heterogeneous study population. In practice, the study population may consist of multiple internally homogeneous subpopulations or latent classes, which complicates the measurement of the latent variable.

The object is to study the relationship of such a continuous latent variable to survival where the study population is heterogeneous. Relationships can be studied between some time to an event (e.g. death) and changes in the latent variable. Typically, to study the relationship of a latent variable to survival, longitudinal measurements are required to be able to identify a patient's latent variable trajectory. The

^aTNO Quality and Safety, Zeist, The Netherlands

^bDepartment of Research Methodology, Measurement, and Data Analysis, Faculty of Behavioral Sciences, University of Twente, Enschede, The Netherlands

^cMRC Biostatistics Unit, Institute of Public Health, Cambridge, U.K.

*Correspondence to: Jean-Paul Fox, Department of Research Methodology, Measurement, and Data Analysis, Faculty of Behavioral Sciences, University of Twente, Enschede, The Netherlands.

†E-mail: j.p.fox@utwente.nl

latent variable trajectory provides information about the development of the individual's behavior or functioning and can also be used to understand how specific changes may be related to covariates [5, 6]. In case the latent variable trajectories are nested within different subgroups or latent classes, it is also possible to evaluate group-specific behavior or functioning over time and group differences with respect to survival.

Item response theory (IRT) models [7] have been developed to take account for measurement error inherent to a questionnaire, mixed response types and different item characteristics, among other things. Specifically, the generalized partial credit model [8, 9] will be used to establish a functional relationship between the observed discrete item responses and the continuous latent variable. The measurement occasions over time are nested within the individual. Therefore, a multilevel generalized partial credit model is proposed to model the individuals' latent developmental trajectories, which relates to the multilevel IRT model approach of Fox and Glas [10].

Heterogeneity within the population as a result of the presence of unidentified subgroups can be dealt with by a mixture modeling approach [11, 12]. Mixture models can identify groups of persons that are relatively homogenous in their developmental trajectories and assign probabilities of group membership to each person. Moreover, a mixture model provides the means for linking group membership probability to relevant disease-related characteristics [13, 14]. A mixture multilevel IRT model is defined to model the latent developmental trajectories of the measures such that dependencies between measures within the same person are accounted for [15, Chapter 6]. As a result, besides allowing for a deviation in the mean trend of the developmental trajectories of subgroups, the mixture multilevel IRT model also allows for the stratification of the survival function on group membership. Also, by assigning group membership probabilities, this avoids the arbitrary classification of persons to a certain group, which can be important when differences between groups are such that they require different treatments. Vermunt [16] proposed multilevel latent class models to identify subtypes of related cases accounting for a nested structure of the data.

The joint modeling of a latent variable and survival is usually done by extending the Cox proportional hazards regression model with a time-dependent latent covariate [17–19]. Larsen [1] and Wang *et al.* [2] integrated an IRT model with Cox proportional hazard model. The hazard function can take on different forms but the hazard functions of different individuals are restricted to be proportional over time. This assumption is central to the proper estimation and interpretation of the survival model but very difficult to validate [20]. In the proposed modeling framework, a general class of parametric survival models that allow for non-proportional hazards will be adopted. A time-dependent latent covariate is accommodated using the subject-specific latent variable trajectory. Instead of using measurements at fixed time points, the time intervals between the measurements of individuals are left unspecified. As a result, the exact times of measurement and death can be used.

Other authors have proposed joint models for longitudinal and time-to-event data; an overview of this active research area can be found in [21]. However, our proposal builds on previous work by extending the modeling framework with a mixture multilevel IRT model, and the use of Bayesian methods for joint estimation, which are implemented using Markov chain Monte Carlo (MCMC) routines, thereby accounting for measurement uncertainty in the latent covariates that is immediately reflected in the parameter estimates of the survival function [22].

The joint model is motivated by the following example, which comprehends a follow-up study of the development of cognitive impairment and survival of elderly participants. On several measurement occasions, cognitive function of the participants was assessed with the mini-mental state examination (MMSE), a questionnaire widely used for screening of cognitive impairment and used to detect cognitive decline that is typical for people suffering from dementia or cognitive impairment seen by Alzheimer's disease [23]. Typically, besides individual differences, the study population consists of subtypes since a group of participants will show cognitive decline over time as normally seen by elderly people and another group will show a more severe cognitive decline due to the development of Alzheimer's disease or dementia. Furthermore, it is reasonable to suspect different life expectancies for groups with different trajectories of cognitive impairment.

In the next section, the modeling framework is described in more detail. Section 3 discusses the MCMC algorithm. Section 4 introduces the MMSE and survival data set and describes the results and conclusions of the application of our model to the data. A discussion of the merits and possible improvements and extensions of our approach concludes this paper.

2. Modeling latent developmental trajectories and survival

The model is composed of a mixture multilevel item response model and a survival model. The mixture model identifies subpopulations that are relatively homogenous in their developmental trajectories. Within each subpopulation, a multilevel latent growth model is used to describe the individuals' latent developmental trajectories. Subsequently, an item response model maps an individual's response vector on several indicator variables to a test score on the latent trait. A parametric survival model is stratified on group membership, and links (latent) covariates to the hazard of an event.

2.1. Specification of the mixture multilevel item response model

Assume there are $g = 1, \dots, G$ unobserved groups in the population and that subjects ($i = 1, \dots, N$) belong to one of the subgroups. For each subject i , a latent variable θ_{ij} is measured at occasion j at time t_{ij} . The number of measurements can vary across individuals and n_i denotes the number of measurements of subject i . The measures are allowed to be taken at different occasions and the time intervals between measurements are not necessarily equidistant but are allowed to differ across individuals and can, for instance, depend on the measurement frequency of covariates.

Let Ω_g denote the vector of class-specific parameters describing the latent developmental characteristics of group g . The mixing probability is represented by π_g , which is the proportion of individuals belonging to group g . At level 3, the mixture of distributions that describes the population is given by:

$$p(\boldsymbol{\theta} | \boldsymbol{\Omega}) = \sum_{g=1}^G \pi_g p(\boldsymbol{\theta} | \boldsymbol{\Omega}_g) = \sum_{g=1}^G \pi_g \prod_{i=1}^N p(\theta_i | \boldsymbol{\Omega}_g), \quad (1)$$

where it is assumed that the vectors of individual measurements are conditionally independent of one another given the class-specific parameters. The conditional probability that subject i belongs to group g is given by:

$$P(Q_i = g | \boldsymbol{\theta}_i, \boldsymbol{\pi}, \boldsymbol{\Omega}) = \frac{\pi_g p(\boldsymbol{\theta}_i | \boldsymbol{\Omega}_g)}{\sum_{g=1}^G \pi_g p(\boldsymbol{\theta}_i | \boldsymbol{\Omega}_g)}, \quad (2)$$

where Q_i is the underlying discrete latent class variable that can take on values $g = 1, \dots, G$.

At level 2, a latent curve model is specified to describe the latent developmental trajectories. The common form of the latent curve model is one based on linear change. However, within the joint modeling framework, a more general polynomial (such as linear or quadratic) or nonlinear growth model is also possible. To account for the nesting of measures within subjects in latent class g , a linear growth model with a random intercept and slope at the individual level is specified by

$$\theta_{ij} = \gamma_{0,g} + u_{0i,g} + t_{ij}(\gamma_{1,g} + u_{1i,g}) + \mathbf{w}_{i,g} \boldsymbol{\gamma}_{2,g} + e_{ij}, \quad (3)$$

where $\gamma_{0,g}$ denotes the class-specific intercept, and $\gamma_{1,g}$ denotes the class-specific slope or the annual change rate. The individual-level coefficients $u_{0i,g}$ and $u_{1i,g}$ are the person-specific random effects that represent the deviation from the class-specific intercept and slope, respectively. The set of random effects is assumed to be multivariate normally distributed as $(u_{0i,g}, u_{1i,g}) \sim \text{MVN}(\mathbf{0}, \boldsymbol{\tau}_g)$. The subject and class-specific explanatory information are stored in matrix $\mathbf{w}_{i,g}$, and $\boldsymbol{\gamma}_{2,g}$ represent the corresponding (fixed) effects. The errors, e_{ij} , are independently and normally distributed with mean zero and variance σ^2 .

At level 1, the generalized partial credit model [8, 9] is adopted. It relates the latent variable θ_{ij} to K observed responses $\mathbf{Y}_{ij} = (Y_{ij1}, \dots, Y_{ijK})$, and the probability of a response of subject i on item k in category c ($c = 1, \dots, C_k$) is defined as

$$P(Y_{ijk} = c | \theta_{ij}, a_k, \mathbf{b}_k) = \frac{\exp\left[\sum_{v=1}^c a_k(\theta_{ij} - b_{kv})\right]}{\sum_{h=1}^{C_k} \exp\left[\sum_{v=1}^h a_k(\theta_{ij} - b_{kv})\right]}, \quad (4)$$

where a_k is the slope parameter and $\mathbf{b}_k = (b_{k1}, \dots, b_{kC_k})$ are the threshold parameters of item k . The number of response categories per item may differ. To identify the categories for each item, the threshold or step parameter b_{k1} is fixed at zero.

For polytomous item responses, each threshold parameter in equation (4) is associated with a response category and defines the difficulty of completing that step or getting a score in that category. The numerator contains the steps that are completed and the denominator is the sum of all possible numerator terms.

For binary item responses, the only threshold parameter b_k is the item's difficulty parameter. Then, the probability of a correct response can be stated as

$$P(Y_{ijk} = 1 | \theta_{ij}, a_k, b_k) = \frac{\exp[a_k(\theta_{ij} - b_k)]}{1 + \exp[a_k(\theta_{ij} - b_k)]}, \quad (5)$$

where $P(Y_{ijk} = 0 | \theta_{ij}, a_k, b_k) = 1 - P(Y_{ijk} = 1 | \theta_{ij}, a_k, b_k)$. The difficulty parameter defines a location on the scale of the latent variable where the probability of a correct response equals 0.50. When increasing the item difficulty, a similar increase in the latent-variable value is required to have a constant success probability.

The slope parameter in the generalized partial credit model is also referred to as the discrimination parameter and is interpreted as a characterization of an item. It influences the increase in probability of scoring in a response category when increasing the level of the latent variable. An item discriminates well when persons with different latent-variable levels have different probabilities of scoring in the same category, where the person with the higher latent-variable level has a higher chance.

If the items in the test measure one latent variable, it can be assumed that the person's item responses are independent given the value of latent variable θ_{ij} . That is, the latent variable explains all associations between the item responses. In the generalized partial credit model it is assumed that such a unidimensional latent variable underlies the item responses. As a result, the joint density of the K item responses of subject i given the latent variable and item parameters can be factorized as

$$p(y_{ij1}, \dots, y_{ijK} | \theta_{ij}, \mathbf{a}, \mathbf{b}) = \prod_{k=1}^K p(y_{ijk} | \theta_{ij}, a_k, \mathbf{b}_k). \quad (6)$$

2.2. Specification of the parametric survival model

The observed continuous survival time data, \mathbf{t} , present the time to a certain event. The survival times are right-censored for subjects that are still alive at the end of the study. An indicator variable is defined to identify the uncensored times and the right-censored times. Let $D_i = 1$ when subject i experienced an event before the end of the study, and $D_i = 0$ when the event did not occur and a right-censored survival time is observed.

The distribution of the survival times of subjects in latent class g are characterized by survival function $S_g(t)$. It represents the conditional probability that subject i in latent class g does not experience an event before time t_i . The latent class-specific survivor function depends on a location parameter η_g and some shape or scale parameters $\boldsymbol{\alpha}$. The distribution can be any well-known parametric survival distribution (e.g. exponential, Weibull, lognormal). Let $f_g(t)$ denote the density function of the observed survival time, and the survivor function is related to the cumulative distribution function according to $S_g(t) = 1 - F_g(t)$.

Consider the time interval $[t_{i0}, t_{i1}]$ where subject i in class g enters the study at time t_{i0} and is followed up to time t_{i1} . The distribution of the time t_{i1} is described by the density function $f_g(t)$ when an event has occurred in the time interval or by the survivor function when subject i is still at risk; that is,

$$p(t_{i1}, d_{i1} | \eta_g, \boldsymbol{\alpha}, Q_i = g) = \frac{S_g(t_{i1} | \eta_g, \boldsymbol{\alpha})^{1-d_{i1}} f_g(t_{i1} | \eta_g, \boldsymbol{\alpha})^{d_{i1}}}{S_g(t_{i0} | \eta_g, \boldsymbol{\alpha})}. \quad (7)$$

Note that the denominator accounts for possible late entry of subject i , and $S_g(t_{i0}) \equiv 1$ at the start of the study to simplify the expression.

The survival model is extended to identify risk factors and to allow class-specific individual differences in survival. Therefore, a loglinear model with class-specific parameters is defined for the location parameter of the survivor function, which is given by

$$\log \eta_{ij,g} = \mathbf{x}_{ij}^t \boldsymbol{\beta}_g + \theta_{ij} \Lambda_g, \quad (8)$$

where \mathbf{x}_{ij} is the vector of corresponding time-dependent observed covariates, $\boldsymbol{\beta}_g$ the vector of corresponding class-specific effects, and Λ_g is the class-specific regression weight for the time-dependent

latent variable θ_{ij} . The latent variable is a time-dependent covariate in equation (8), which is measured infrequently and with measurement error. However, the joint modeling framework for the latent covariate and survival process can deal with the random error in the measured covariates. Various modeling approaches for dealing with a latent covariate as a predictor have been proposed [19, 24, 25].

The time-dependent covariates in equation (8) enter the survivor function as time-varying covariates, which severely complicate the evaluation of the survivor function since it requires integrating the hazard function over each covariate trajectory [26, pp. 357–358]. Therefore, it will be assumed that the time-dependent covariates are piecewise-constant within a specified interval. Such a finite partition of the time axis is a convenient and popular method for semiparametric survival analysis.

For subject i , n_i intervals $(t_{i0}, t_{i1}), (t_{i1}, t_{i2}), \dots, (t_{i(n_i-1)}, t_{in_i})$ are defined, where each time point reflects a new measurement occasion of the time-dependent covariate(s). Now, the distribution of the survival times \mathbf{t}_i given time-dependent covariates is described by

$$\begin{aligned} p(\mathbf{t}_i, \mathbf{d}_i | \boldsymbol{\eta}_i, \boldsymbol{\alpha}) &= \sum_{g=1}^G P(Q_i = g) \prod_{j=1}^{n_i} p(t_{ij}, d_{ij} | \eta_{ij,g}, \boldsymbol{\alpha}, Q_i = g) \\ &= \sum_{g=1}^G P(Q_i = g) \prod_{j=1}^{n_i} \frac{S(t_{ij} | \eta_{ij,g}, \boldsymbol{\alpha})^{1-d_{ij}} f(t_{ij} | \eta_{ij,g}, \boldsymbol{\alpha})^{d_{ij}}}{S(t_{i(j-1)} | \eta_{ij,g}, \boldsymbol{\alpha})}. \end{aligned} \quad (9)$$

2.3. Joint likelihood function

The joint likelihood function is derived for the survival model and the mixture multilevel item response model. The first part of the likelihood is constructed from the distribution of the survival data using the expression in equation (9). The second part of the likelihood is constructed from the distribution of the item response data and the latent variable using the expressions in equations (1) and (6). Due to the factorizations, the likelihood for the joint model can be written as a series of products:

$$\begin{aligned} p(\mathbf{y}, \mathbf{t}, \mathbf{d} | \boldsymbol{\Omega}, \mathbf{a}, \mathbf{b}, \boldsymbol{\beta}) &= \prod_{i=1}^N \left[\sum_{g=1}^G P(Q_i = g) \prod_{j=1}^{n_i} p(t_{ij}, d_{ij} | \eta_{ij,g}, \boldsymbol{\alpha}, Q_i = g) \right] \\ &\quad \times \prod_{j=1}^{n_i} \int p(\mathbf{y}_{ij} | \theta_{ij}, \mathbf{a}, \mathbf{b}) \sum_{g=1}^G P(Q_i = g) p(\theta_{ij} | \boldsymbol{\Omega}_g, Q_i = g) d\theta_{ij} \\ &= \prod_{i=1}^N p(\mathbf{t}_i, \mathbf{d}_i | \boldsymbol{\eta}_i, \boldsymbol{\alpha}) p(\mathbf{y}_i | \boldsymbol{\Omega}, \mathbf{a}, \mathbf{b}). \end{aligned} \quad (10)$$

The first term on the right-hand side denotes the likelihood contribution of the survival component. A heterogeneous population distribution is defined using a mixture model such that different survivor functions can be defined across subpopulations. The survivor functions vary across individuals within the same class through time-dependent (latent) covariates. The second term on the right-hand side denotes the distribution of the item responses given the individual-specific latent trajectory parameters and item parameters. The characteristics of the latent developmental trajectories are assumed to vary across the subpopulations, and individual variability in the same class is captured through random effects.

In this joint likelihood model, interest is focused on how the latent variable trajectory affects the time to some event. In fact, the questionnaire data are assumed to be informative about the survival data using a continuous latent variable and a discrete latent class as predictors. Besides the item response data, the survival data can be used to define the latent class membership. Then, the survival data are informative about the latent variable trajectory through their influence on the specification of the latent classes. Larsen [27] recommended to define the latent variable trajectory by the item response data and not the survival data. This is possible by using the item response data and not the survival data to define the latent classes. Hogan and Laird [28] defined the joint likelihood model in such a way that the time to some event, such as informative study dropout, affects a longitudinal outcome. In the present framework this would mean that the survival data, and not the item response data, define the latent classes.

2.4. Model identification

To ensure identification of the measurement model for θ , the scale of the latent variable has to be fixed. This is achieved by specifying the general mean of the latent variable to be zero and the $\prod_{k=1}^K a_k = 1$, which fixes the location and scale, respectively. Furthermore, the mixture model is yet unidentified and requires a restriction to be able to distinguish between the G permutations of the mixture components (label switching). A unique solution requires either a restriction on the means, the variances or the mixture proportions [29]. In our implementation the mixture model was identified by the restriction $\tau_1 < \tau_2 < \dots < \tau_G$. For two groups, for example, this was accomplished by restricting $\tau_2 - \tau_1$ to be positive definite.

3. Bayesian estimation and inference

Posterior inferences require the specification of prior distributions for all model parameters. The prior is then updated with the likelihood given by (10) to obtain the posterior. This section discusses the choices for the priors and outlines the MCMC algorithm for estimation. Furthermore, methods for model evaluation are discussed.

3.1. Estimation

Within the mixture multilevel IRT model, the hierarchical structure on the latent variable naturally provides priors for the parameters. Priors need to be specified for the higher-level model parameters and some level-specific model parameters. First, at the level of response observations, an exchangeable hierarchical multivariate normal prior is chosen for the item parameters of the measurement model such that $\log(a_k), b_k \sim \text{MVN}(\boldsymbol{\mu}_I, \boldsymbol{\Sigma}_I)$. It will be assumed that there is no prior information to discriminate the item parameters from one another. A possible within-item dependency is modeled by the multivariate prior structure to allow correlation between item's discrimination and difficulty parameter. A normal-inverse Wishart prior will be used for the hyperprior parameters.

Second, priors are defined for the mixture multilevel model parameters. Normal priors are chosen for the fixed effects parameters, allowing for efficient Gibbs sampling steps for estimation (see below), $\boldsymbol{\gamma} \sim \text{MVN}(\boldsymbol{\mu}_{\gamma_0}, \boldsymbol{\Sigma}_{\gamma_0})$. A conjugate prior is specified for the mixture proportions π_g ; that is, the beta distribution, in the case of two classes, and the Dirichlet distribution, in the case of more than two latent classes. For the variance components, an inverse-gamma prior is chosen for σ^2 and its multivariate generalization, the inverse-Wishart distribution, for $\boldsymbol{\tau}_g, \boldsymbol{\tau}_g \sim IW_{v_\gamma}(\boldsymbol{\tau}_0^{-1})$, with v_γ the prior degrees of freedom and scale $\boldsymbol{\tau}_0$.

Third, for the survival model, prior distributions for the regression and shape parameters have to be specified. The regression parameters, $\boldsymbol{\beta}, \boldsymbol{\Lambda}$, are given independent normal priors that reflect only a non-informative region of possible parameter values. The specific choices for the mean and variance are given below. For the shape or scale parameters uniform priors are defined on a finite interval with sufficiently large bounds and independent of the choice of the survival function.

To obtain parameter estimates, an MCMC method was implemented. MCMC methods are simulation-based algorithms that construct a Markov chain with the joint posterior distribution as its target distribution. The joint distribution is broken down into smaller subvectors, conditional on all other parameters. After providing the algorithm with starting values for all parameters, it alternates between the full conditional distributions for M iterations. The details of the algorithm are given in the Appendix, but we illustrate the procedure briefly here:

1. Generate starting values for all parameters
2. At iteration m , draw latent variable $\boldsymbol{\theta}^{(m)}$ from $p(\boldsymbol{\theta} | \mathbf{y}, \mathbf{t}, \mathbf{w}, \mathbf{a}, \mathbf{b}, (\text{all other parameters}))$
3. Draw survival parameters $\boldsymbol{\Lambda}$ from $p(\boldsymbol{\Lambda} | \boldsymbol{\theta}^{(m)}, \mathbf{y}, \mathbf{t}, \mathbf{w}, \mathbf{a}, \mathbf{b}, (\text{all other parameters}))$
4. Obtain draws for all other model parameters
5. Repeat steps 2–4 until M draws from the joint posterior have been obtained.

After M draws have been obtained, it is important to assess the convergence of the algorithm before proceeding with posterior inferences. The BOA R-package [30] for use in R/Spplus allows to evaluate several statistical tests that give an indication of the convergence of the MCMC chains. A burn-in period for the chains will be estimated and the corresponding draws are ignored for making statistical

inferences. The remaining draws are used to compute summary statistics of the posterior distribution of the model parameters.

Note that at iteration m , a draw of the latent variable $\theta^{(m)}$ from the posterior distribution is treated as a covariate value in the individual survival function. This is done at every iteration. This way measurement error in the latent covariate is also accounted for in the estimated posterior density of the survival function parameters. The MCMC algorithm makes it possible to estimate the marginal posterior densities of the survival parameters by integrating over the posterior density of the latent explanatory variable.

3.2. Model evaluation

The framework of posterior predictive testing is used to investigate the fit of the model [31, 32]. The principle is to evaluate general model fit or to test a specific assumption using an appropriate test statistic. The value of the test statistic is computed given the observed data and its posterior distribution is computed given replicated or predictive data sampled from their posterior predictive distribution under the model. The extremeness of the value of the test statistic is evaluated, and a misfit is detected when the value is more extreme than expected under the model. Values of the test statistics and draws from the posterior predictive distribution can be obtained as by-products of the MCMC algorithm. Posterior predictive checks for item response models have been proposed by Sinharay [33]. Fox [15] provides a general overview of prior and posterior predictive checks for item response models.

Posterior predictive model assessment can also be applied to the general class of survival models. Ibrahim *et al.* [26] and Hanson *et al.* [34] used predictive model selection criteria to compare models. The posterior predictive survival data can be sampled under the joint model using the MCMC samples. Here, model fit is assessed by evaluating the quality of the predictions of the time-to-event data under the joint model. Replications of the event indicator, \mathbf{D}^{rep} , are sampled under the joint model. Then, we define O_t as the cumulative number of events after t years since the start of the study at $t=0$, and $E(O_t)$ as its expectation under the model. The corresponding model fit criterion is defined as

$$T(\mathbf{D}^{\text{rep}}, \mathbf{t}, \mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{\beta}) = \sum_t \frac{(O_t^{\text{rep}} - E(O_t))^2}{E(O_t)}. \quad (11)$$

The observed cumulative number of events, \mathbf{O}^{obs} , are considered to be extreme under the joint model, indicating a model misfit, when the posterior probability $P(T(\mathbf{D}^{\text{rep}}, \mathbf{t}, \mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{\beta}) > T(\mathbf{D}^{\text{obs}}, \mathbf{t}, \mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{\beta}))$ is close to zero or one.

Joint models can differ with respect to their survival component and/or the longitudinal component. The deviance information criterion (DIC) allows to judge the performance of different (non-nested) models by means of a deviance-based statistic with a penalty term for the number of parameters in the model [35]. The specification of the deviance function defines the type of model comparison that will be performed. Here, the DIC is based on the likelihood of the observed data, where the parameters of interest can be directly identified and the DIC can be computed in a closed form. Therefore, a DIC can be defined for the observed item responses and for the observed survival times.

When interest is focused on selecting the appropriate survivor function, the deviance function will be defined as $-2 \log p(\mathbf{t}, \mathbf{d} | \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\theta})$. Different parametric survival models (e.g. lognormal, Weibull) can be compared, and specific shape parameters for the survivor functions and the effects of explanatory variables can be tested. This DIC enables the selection of a parametric survivor function for the observed survival times. For the item response data, the DIC can also be defined at the level of the observed data and will be a function of the observed-data likelihood defined by the generalized partial credit model. Both DICs can be estimated using the MCMC samples and the closed-form expressions of the deviance functions.

4. Data analysis

Data were obtained from a study of the OPTIMA cohort (Oxford Project to Investigate Memory and Ageing), which were supported by the NIHR Oxford Comprehensive Biomedical Research Centre. The sample consisted of 668 participants, aged 65 and older, who were followed over the period 1991–2003. Participants were interviewed using the MMSE questionnaire to obtain a measure of cognitive function. Furthermore, background variables such as gender, age, education and a known family history

of dementia were recorded. Follow-up times varied between as well as within individuals and in total 3668 measurement occasions were available, corresponding to the total number of interviews, right-censored states and observed deaths. The minimum number of observations per individual was 2, the maximum 16. The median and standard deviation of the number of individual measurements was 5 and 2.54, respectively.

The data suggested that the participants could be divided into roughly two groups. On the one hand there were individuals who showed relatively stable test scores over time and who were in the higher levels of the sum score distribution; on the other hand, there were individuals who showed a decline in their test scores over time. Therefore, it was assumed that the population consisted of a subpopulation of participants with latent trajectories showing a decrease in cognitive function and a subpopulation showing more constant trajectories of cognitive function over time. The latent trajectories as well as the survival functions were expected to differ across subpopulations and participants. The objective was to investigate the structure of the population distribution of the latent trajectories, the relationship of cognitive function with survival and to estimate individual survival functions given the survival data and the individual latent trajectories of cognitive impairment.

4.1. The mixture multilevel modeling part

The MMSE item response data were modeled by the generalized partial credit model, where the latent variable represent cognitive function. A linear growth model for the cognitive function was specified with a random intercept and slope parameter for each subpopulation. The follow-up time was used as an explanatory variable. It was assumed that the residual variation between measurements was constant across subjects and subpopulations. This two-component multilevel population model for cognitive function can be stated as,

$$\begin{aligned}
 p(\theta_{ij} | \mathbf{t}_i, \gamma, \boldsymbol{\tau}, \sigma^2) &= \pi_1 \phi(\mu_{ij,1}, \sigma^2) + (1 - \pi_1) \phi(\mu_{ij,2}, \sigma^2), \\
 \mu_{ij,g} &= \gamma_{0i,g} + \gamma_{1i,g} t_{ij}, \\
 \gamma_{0i,g} &= \gamma_{0,g} + u_{0i,g}, \\
 \gamma_{1i,g} &= \gamma_{1,g} + u_{1i,g},
 \end{aligned} \tag{12}$$

for $g = 1, 2$. The population of cognitive impairment is a mixture of two populations. The trajectory function in each subpopulation has its class-specific intercept, slope and population variance. Furthermore, the random intercept and slope models the individual variation within each class. The residuals ($u_{0i,g}$ and $u_{1i,g}$) are multivariate normally distributed, and they reflect the individual deviations from the class-specific mean intercept and slope, respectively.

The MCMC algorithm was run twice for 15 000 iterations, after which convergence was checked using the BOA *R*-package. The two separate runs were compared using Gelman's *R*-statistic, that compares the two chains with respect to their target distribution. If the posterior distributions have converged to approximately the same distribution, this statistic is close to one. Values below 1.1 are deemed acceptable and in our sample all estimated *R* statistics were less than 1.07. Furthermore, Geweke's *Z*-statistic, Heidelberg's stationarity test as well as trace plots and estimated autocorrelations were evaluated. Results indicated that stability of the chains for all parameters was reached within the first 1500 iterations. We decided to discard the first 5000 iterations as burn-in, and based our posterior inferences on the remaining 10 000 samples. This applies to all estimates reported below.

Model fit of the multilevel item response model part was assessed by evaluating an observed score statistic, which gives an impression of overall model fit by comparing the observed sum scores of the participants on the MMSE with their replicated sum scores under the model [33, 36]. The results of 500 replicated data sets under the model showed a good fit for the response data. The local independence assumption in equation (6) was evaluated using a posterior predictive check based on residual correlations. The test results showed that less than 5 per cent of the item responses had non-zero (within-subject) correlations. Therefore, it was concluded that a unidimensional latent variable explained the dependencies between the response observations.

There was no significant variation detected between the individual slopes of the trajectory function. Therefore, the random slope was restricted to be constant across individuals within each subpopulation. The parameter estimates of the mixture multilevel population model (12) are presented in Table I. The two-group solution is given the labels 'Decline' and 'Stable', to refer to the group of subjects who showed

Table I. Estimates of the class-specific developmental trajectories of cognitive function.

	Decline ($g = 2$)		Stable ($g = 1$)	
	EAP	SD	EAP	SD
<i>Fixed effects</i>				
$\gamma_{0,g}$ Intercept	-0.709	0.031	0.776	0.036
$\gamma_{1,g}$ Time slope	-0.112	0.012	-0.009	0.006
<i>Variance components</i>				
τ_g^2 Between individual	0.244	0.124	0.134	0.118
σ^2 Residual	0.219	0.029	0.219	0.029
<i>Mixture proportion</i>				
π_g	0.432	0.020	0.568	0.020
DIC	59 866 ($p_D = 3,452$)			

a decrease in cognitive function over time and those who showed stable performance, respectively. For a formal comparison of the two-group mixture solution with the homogenous population assumption, the DIC was estimated at the level of the observed item responses. The estimated DIC corresponding with the mixture multilevel population model (equation (12)) was 59 866 ($p_D = 3452$), while for the homogenous population model the DIC was 60 052 ($p_D = 3207$), favoring the two-component mixture population model. A three-component mixture model was also fitted, but this led to a higher DIC of 67 235. The three-group solution showed a comparable stable group, where the other two groups shared the characteristics of the decline group. This three-group solution did not provide any additional information.

The estimated mixing population proportion $\hat{\pi}_2$ presents the percentage of subjects who were classified into the decline group. It appeared that around 43 per cent of the subjects belong to the decline group. Note that the probability of membership to a specific class varies over individuals. The individual posterior classification probabilities are not given here but they were estimated with the other model parameters.

The expected *a posteriori* (EAP) estimates presented in Table I show that the mean intercept of the stable group ($EAP(\gamma_{0,1}) = 0.776$) was substantially higher than the mean intercept of the decline group ($EAP(\gamma_{0,2}) = -0.709$). A 95 per cent HPD interval for the difference in mean intercepts was estimated at [1.430, 1.550], and clearly excludes the point of equal mean intercept values. The mean trend in cognitive function over time was almost zero for the stable group, whereas it was negative for the decline group, who showed a downward trend ($EAP(\gamma_{1,2}) = -0.112$). An illustration of the differences between trajectories of cognitive function of the decline and the stable group is given in Figure 1. It shows the developmental trajectories of ten selected subjects. The subjects were classified with a membership probability higher than 0.975. Besides the individual variation, it can be seen that the class-specific mean latent trajectories differ substantially.

4.2. Including the survival modeling part

The modeling structure of the latent growth model for cognitive impairment was identified independent of the survival component. In a second step, the survival component of the joint model was analyzed given the growth modeling structure. It was investigated whether the identified latent groups (e.g. stable and decline) also differ with respect to their survivor functions. Furthermore, the effect of cognitive function, age, and gender (male = 1) on survival was investigated.

Different functional forms of the survivor function were considered; the exponential, Weibull and the lognormal distribution, which are commonly used in survival modeling [37]. The survival models were jointly estimated and the mixture multilevel model and the survivor function was stratified on group membership. This means that (class-specific) survival parameters were estimated for both subpopulations. The survival functions were extended with the individual time-dependent latent and observed covariate θ_{ij} and Age_{ij} , respectively, and the time-invariant covariate gender. This leads to the following subject-specific predictor in the survivor function,

$$\log \eta_{ij,g} = \beta_{0,g} + \text{Gender}_i \beta_{1,g} + \text{Age}_{ij} \beta_{2,g} + \theta_{ij} \Lambda_g. \tag{13}$$

The fit of the different parametric survival distributions was evaluated using the DIC. Table II shows the estimated DICs of the different joint models where the survival component is altered by choosing a different survival function (column labeled 'Survival Distribution'), different covariates (column labeled

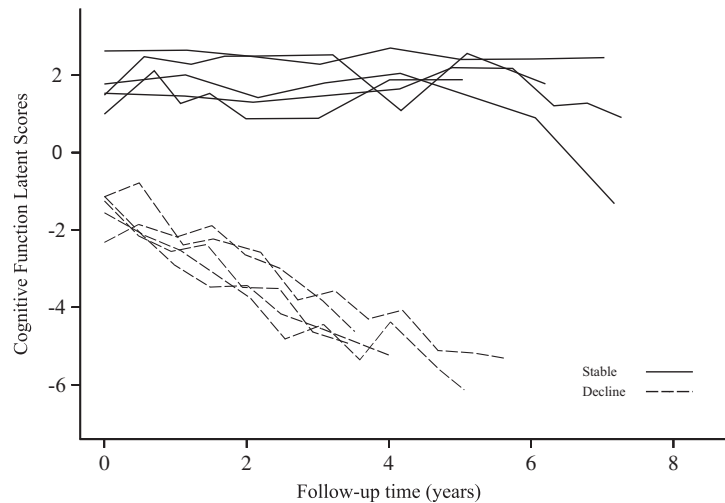


Figure 1. Cognitive developmental trajectories of five subjects in the decline group and five in the stable group.

Table II. Estimated DICs of different stratified and non-stratified survival models.				
Model	Survival distribution	Covariates	Latent classes	DIC
M_1	Exponential	1	2	2146.0
M_2	Weibull	1	2	1872.3
M_3	Lognormal	1	2	1905.4
M_4	Exponential	$1, \theta$	2	2041.7
M_5	Weibull	$1, \theta$	2	1836.0
M_6	Lognormal	$1, \theta$	2	1816.1
M_7	Weibull	$1, \theta, \text{Male}$	2	1823.2
M_8	Lognormal	$1, \theta, \text{Male}$	2	1803.0
M_9	Weibull	$1, \theta, \text{Male, Age}$	2	1775.0
M_{10}	Lognormal	$1, \theta, \text{Male, Age}$	2	1768.1
M_{11}	Weibull	$1, \theta, \text{Male, Age}$	1	1856.3
M_{12}	Lognormal	$1, \theta, \text{Male, Age}$	1	1858.3

θ represents the latent covariate cognitive function.

‘Covariates’) and different number of mixture distributions (column labeled ‘Latent Classes’). It can be seen that the exponential survival distribution (model M_1 and M_4) is too restrictive compared to the Weibull and lognormal. The Weibull and lognormal distributions are very close in fit as indicated by the DIC, but the latter is doing better when gender and cognitive function are included (model M_9 and M_{10}). Finally, the stratified survival models were compared with the non-stratified survival models (models M_{11} and M_{12}), where for the latter the effects across the two subpopulations were assumed to be equal. From the estimated DICs follow that the characteristics of the survival function vary over the subpopulations.

The covariates were entered sequentially into the model and the DIC as well as 95 per cent HPD regions of the estimated coefficients were used to test the statistical significance of the covariates. From Table II it follows that the smallest DIC corresponds to model M_{10} . The model fit was assessed by means of a posterior predictive check. For the proposed posterior predictive check of the survival part (equation (11)), the observed cumulative number of deaths per year since baseline (1, 21, 67, 108, 143, 198, 227, 251, 266, 286) were compared to the replications under the model. The estimated p-value was $p=0.37$, which indicates that the observed number of deaths cannot be considered to be extreme under the model, although the model seemed to slightly underpredict the number of observed events.

The parameter estimates of the survival component (model M_{10}) are given in Table III. The parameter estimates are given for the subpopulations labeled Stable and Decline. The residual variance was restricted to be the same for both groups. The survival probability is lower for the decline group when comparing it with the stable group since for the last mentioned the intercept value is higher. A

Table III. Class-specific estimates of the (stratified) lognormal survival model (M_{10}).

	Decline ($g=2$)		Stable ($g=1$)	
	EAP	SD	EAP	SD
<i>Fixed effects</i>				
$\beta_{0,g}$ Intercept	1.986	0.072	2.561	0.081
$\beta_{1,g}$ Male	-0.270	0.064	-0.282	0.076
$\beta_{2,g}$ Age (standardized)	-0.179	0.080	-0.212	0.090
Λ_g Cognitive Function	0.369	0.041	0.254	0.046
<i>Variance components</i>				
σ_S^2 Residual	0.362	0.031	0.362	0.031

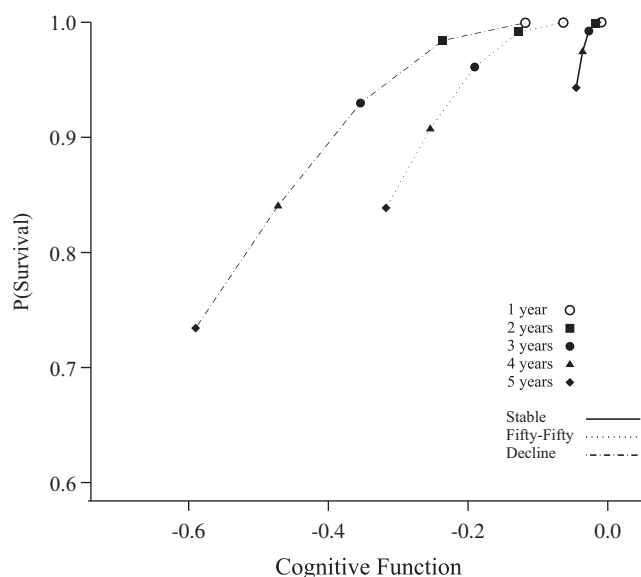


Figure 2. The survival probabilities of three females as a function of a five-year trend in cognitive function.

subject's survival probability is positively correlated with their (grand-mean centered) cognitive function measured at that same time-point. Therefore, a change in the within-subject's cognitive function over time leads to a change in the survival probability. For a decrease in cognitive function, the decrease in survival probability is much higher for subjects in the decline group than for subjects in the stable group. The preservation of cognitive function clearly has a positive effect on survival, especially for those that show cognitive impairment. For both groups, the survival probability is decreasing as a function of age, where the effect of age is slightly smaller for the decline group. Furthermore, males have lower life expectancies than females and this effect is almost similar across groups.

The time-dependent explanatory variables age and cognitive function make it possible to estimate individual-specific survival probabilities. A subject's latent trajectory of cognitive function can be used to predict the survival function, which is also influenced by the subpopulation's trend in survival. The joint modeling framework makes it possible to predict cognitive function as well as the survival probabilities. An illustration is given in Figure 2, where the expected survival probabilities are plotted for three different subjects as a function of the expected cognitive function. One female was classified in the decline group, one in the stable group and the other one was assigned a probability of 50 per cent to belong to either group. The predicted trends in cognitive function and corresponding survival probabilities were plotted for a period of five years. The figure illustrates the differences in expected trends between the subjects and subpopulations, both in expected survival probability and in expected cognitive function. Typically, the predicted trend in cognitive function of the subject in the decline group is much steeper than that of the others. The corresponding predicted survival probabilities are also decreasing faster.

The key advantage of using a joint modeling framework in this study was that subjects who are most at risk of developing a major or minor declining trend in cognitive function could be identified.

The integrated mixture modeling approach extracted meaningful subgroups who differ with respect to the relationship between survival and cognitive impairment. The identification of subjects nested in subgroups who are most at risk assists in the development of specific interventions or treatments.

4.2.1. Using sum scores. The individual MMSE item scores can be summed to compute a sum score for each individual. The sum score can be considered to be an estimate of cognitive function and can be directly incorporated in the survivor function in Equation (13). However, the use of sum scores will diminish the power of identifying subject-specific effects and may lead to biased estimates of the corresponding covariate effect on survival.

The sum scores are treated as measured without an error. The measurement error associated with the sum scores can have an effect on the estimated covariate effects [25]. The latent variable estimates have subject-specific measurement errors since the observed item response patterns are not equally likely to be observed. Furthermore, the generalized partial credit model can handle floor and ceiling effects, since the latent variable is defined on a continuous scale, handle missing item responses and deal with item characteristic differences. Finally, some response patterns can be marked as inconsistent or aberrant. This relates to the fact that a subject who has difficulties to pass the easy items will probably fail the more difficult items. When using sum scores, this type of inconsistency cannot be detected and will be ignored.

When using sum scores, the information about the relationship between the test items and cognitive function is lost. That is, many response patterns lead to equivalent sum scores, though some response patterns are more likely to be observed than others. The different response patterns lead to different latent variable estimates under the generalized partial credit model. Therefore, it will be possible to distinguish respondents from each other given the latent variable estimates, though they are related to the same sum score. For this reason, a more profound relationship can be expected between the latent variable scores and survival than the sum scores and survival. That is, the estimated subject-specific survival probabilities will differ when using sum scores due to a different cognitive-function estimate and a different covariate effect on survival.

In Figure 3, the sum scores are plotted against the latent variable estimates under the generalized partial credit model. It can be seen that the discrete sum scores suffer from a floor and ceiling effect and are not normally distributed. Each individual sum score corresponds to numerous different response patterns and different individual latent variable estimates. The latent variable estimates are defined on a continuous scale and show more variation.

The sum scores of the subjects in both latent groups were computed and re-scaled such that the mean and variance of the distribution of the sum scores correspond with those of the latent variable scores. The transformed sum scores were used as an empirical proxy covariate of cognitive function in model M_{10} (equation (13)). The estimated covariate effect is smaller (0.19) for the Stable group and slightly higher for the Decline group (0.40), when comparing them to the covariate effects in Table III. The within-individual change in the sum score is rougher than the change in the latent variable, but the average within-individual change in each subpopulation does not differ much. This lead to relatively small differences in estimated covariate effects. Note that the estimated latent classes were assumed to be known, since the discrete sum scores will violate the normality assumption of the two-component growth model in equation (12).

In Figure 3, individual predictions of survival are plotted for respondents that have the same sum score at the baseline ($t=0$). In the three subplots, the solid line represents the mean predicted survival probabilities of individuals with a sum score of 15, 20 and 25, respectively. In each subplot, the dotted lines correspond to predicted survival probabilities based on latent variable estimates of individuals with the same sum score. The predicted survival probabilities show much more between-subject variation due to differences in latent variable estimates of cognitive function. For example, for a sum score of 25, an expected difference in survival probability of approximately 0.20 was computed after five years keeping other variables constant. The more accurate latent variable scores led to improved individual predictions of survival.

5. Discussion

A mixture multilevel joint model is proposed to model latent developmental trajectory of cognitive function and survival probability. Cognitive function is considered to be a risk factor of survival, where

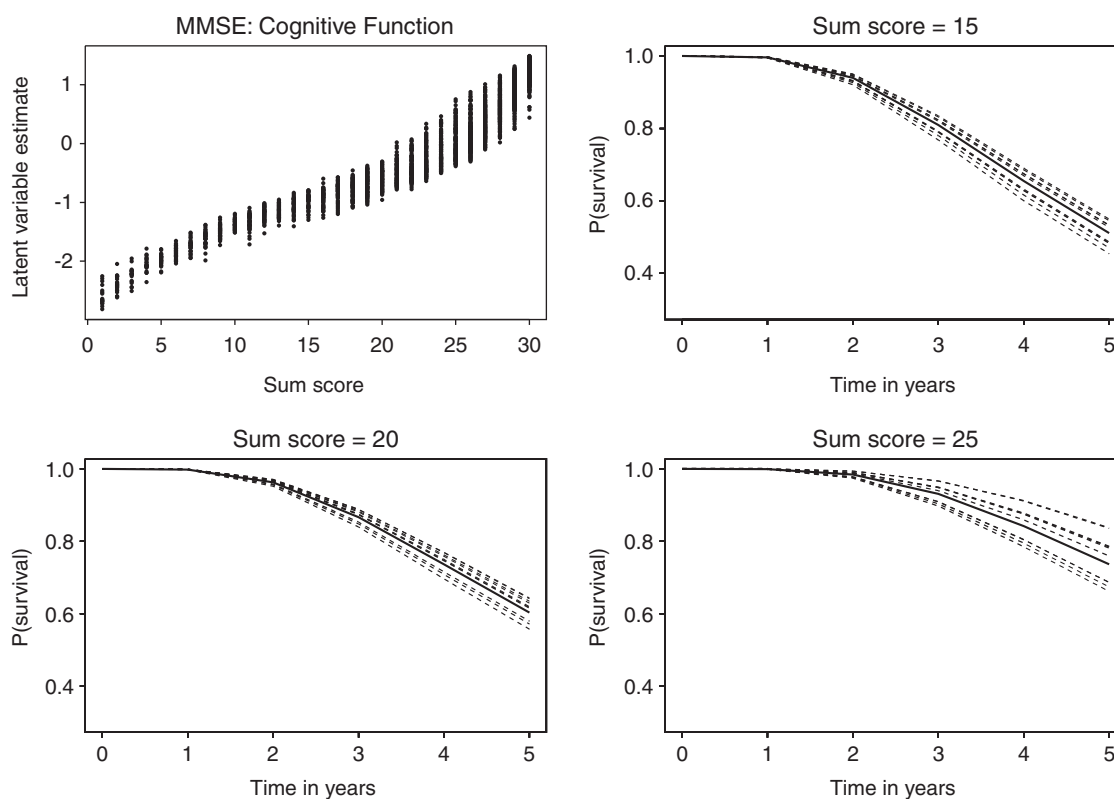


Figure 3. Plot of the sum score against the latent variable estimate of cognitive function. Survival probability plots as a function of time for individuals with different latent variable estimates but an equal sum score.

effects are allowed to vary across subgroups. Longitudinal questionnaire data are used to estimate cognitive function over time. In the joint modeling framework the measurement error in cognitive function is explicitly modeled and the estimated effect of cognitive function on survival does not suffer from attenuation due to the unreliability of the scale scores.

The use of questionnaires by clinicians is widespread (e.g. the measurement of depression or quality of life). Item response theory models become more and more popular for measuring the underlying variable while recognizing the psychometric properties of the measurement instrument being used. The novel proposed mixture item response modeling approach can handle mixed response types as well as an unobserved mixed population of respondents.

In the empirical example it was shown how the proposed joint survival model can be applied to studies where the measures of interest are not directly observable but measured with error by means of a questionnaire. With the proposed model it is possible to identify subgroups within the population and study differences in their growth curves as well as the relationship with time-to-event problems. Neale *et al.* [38] also studied survival and cognitive decline using the MMSE. An advantage of our mixture modeling approach is that the arbitrary classification of individuals based on their MMSE sum scores is avoided. This includes the arguable specification of a cut-off score to identify subjects with conserved cognitive function and impaired cognitive function. In our approach, the uncertainty in individual trends of cognitive impairment is also explicitly modeled over time by means of probability statements.

There is a large literature on joint modeling of (longitudinal) questionnaire data and time-to-event data. However, only a few applications use an item response model for measuring an underlying variable given questionnaire data. Larsen [1] combines a two-parameter item response model with the Cox proportional hazards model. However, a time-independent measurement of physical functioning is defined, which does not facilitate the modeling of a latent developmental trajectory. Wang *et al.* [2] combines an item response model with the Cox proportional hazards model that allows for a longitudinal measurement of a continuous latent variable representing quality of life. They model the mean latent trajectory with a first-order autoregressive structure for the covariance matrix. In the present more general approach, individual latent trajectories of cognitive function are modeled and the time intervals

between the measurements are allowed to vary over individuals instead of assuming fixed follow-up time points. Furthermore, a mixture item response model is used to capture the heterogeneity between meaningful subgroups.

The joint model can also be used for informative dropout in a longitudinal study, where the survival point is taken to be the time to dropout and interest is primarily focused on modeling the latent trajectory. The mixture modeling approach can be used to identify trajectories that are more or less likely to be associated with the dropout. With a focus on the longitudinal modeling part, it is also possible to define a separate (time-dependent) frailty random effect in the survival model, which is correlated with the time-dependent continuous latent variable (e.g. quality of life, cognitive function). As discussed in Wang *et al.* [2], the frailty parameter in the survival model captures all residual individual variation, where the correlation parameter between the frailty parameter and the latent variable captures the dependence in the censoring of the longitudinal process by death.

The statistical inference was performed within the Bayesian framework. An MCMC algorithm is presented to estimate all parameters of the joint model simultaneously. A DIC is implemented to be able to select the (survival) model that best fits the data from a couple of candidate models. The assessment of the joint model is further supported by posterior predictive model checking. However, more research is needed to expand the class of posterior predictive checks to fit joint models. The posterior predictive tests are easily integrated within the MCMC algorithm. Currently, there is no readily available software to fit the proposed joint model. The developed software for the described application will be made available as an R-package from the authors' web site.

Appendix A: MCMC algorithm

Step 1. Sampling classification variable for person i . The full conditional that person i belongs to a specific class is given by (2). Evaluating (2) is straightforward. Subsequently, a classification variable value is drawn given the conditional class-membership probabilities $P(Q_i = g | \theta_i, \pi, \Omega)$ for $g = 1, \dots, G$.

Step 2. Sampling the mixture probabilities. For the two class solution, the number of people in class 1 follows a binomial distribution with N observations. A conjugate and uninformative prior for the class probability π_1 is the Beta(1, 1) and $\pi_2 = 1 - \pi_1$. Let $g_1 = \sum_{i=1}^N I(Q_i = 1)$ and $g_2 = \sum_{i=1}^N I(Q_i = 2)$. The posterior distribution for the latent class probability π_1 is a Beta:

$$\pi_1 | g_1, g_2 \sim \text{Beta}(g_1 + 1, g_2 + 1). \tag{A1}$$

Step 3. Sampling latent variable θ_{ij} . The likelihood is given by (6), multiplying this by the prior $p(\theta_{ij} | \Omega_g, Q_i = g, t_{ij})$ gives the posterior $p(\theta_{ij} | \mathbf{y}_{ij}, \mathbf{a}, \mathbf{b}, \Omega_g, Q_i = g, t_{ij})$. A Metropolis–Hastings step is used to update θ_{ij} . For iteration m , a proposal θ^* is drawn from $J_t(\theta^* | \theta_{ij}^{(m-1)})$, where the jumping distribution J_t was chosen to be $\theta^* \sim N(\theta_{ij}^{(m-1)}, \sigma_{J_t}^2)$. Subsequently, draw $U \sim \text{Uniform}(0, 1)$, and evaluate if $U < r$, where r is the ratio given by

$$r = \frac{p(\theta^* | \mathbf{y}_{ij}, \mathbf{a}, \mathbf{b}, \Omega_g, Q_i = g, t_{ij}) / J_t(\theta^* | \theta_{ij}^{(m-1)})}{p(\theta_{ij}^{(m-1)} | \mathbf{y}_{ij}, \mathbf{a}, \mathbf{b}, \Omega_g, Q_i = g, t_{ij}) / J_t(\theta_{ij}^{(m-1)} | \theta^*)}. \tag{A2}$$

When $U < r$, the proposal is accepted as the new value for θ_{ij} , otherwise the previous draw is retained.

Step 4. Sampling item parameters a_k and \mathbf{b}_k . The likelihood is given by $\prod_{i=1}^N \prod_{j=1}^{n_i} p(y_{ijk} | \theta_{ij}, a_k, \mathbf{b}_k)$. Multiplying this likelihood with the priors for a_k and \mathbf{b}_k gives the posterior. Metropolis–Hastings steps, similar to the one in Step 3, can be defined for a and b . A lognormal proposal distribution is specified for the discrimination parameter.

Step 5. Sampling the fixed effects γ_g . The likelihood is given by (3) and with a normal prior the posterior is multivariate normal. Let $\mathbf{z}_g = (\mathbf{1}, \mathbf{t}_g, \mathbf{w}_g)$ represent the covariate matrix for all persons in class g . Also, let $\boldsymbol{\theta}^* = \boldsymbol{\theta}_g - \mathbf{u}_{0,g} - \mathbf{t}_g \cdot \mathbf{u}_{1,g}$, the vector of person effects for all individuals in class g . Then, the posterior distribution is given by:

$$\gamma_g | \boldsymbol{\theta}, \mathbf{u}, \mathbf{z}_g, \sigma \sim \text{MVN} \left(\frac{\mathbf{z}_g^t \boldsymbol{\theta}^* \sigma^{-2} + \boldsymbol{\Sigma}_{\gamma_0}^{-1} \boldsymbol{\mu}_{\gamma_0}}{(\mathbf{z}_g^t \mathbf{z}_g \sigma^{-2})^{-1} + \boldsymbol{\Sigma}_{\gamma_0}^{-1}}, ((\mathbf{z}_g^t \mathbf{z}_g \sigma^{-2})^{-1} + \boldsymbol{\Sigma}_{\gamma_0}^{-1})^{-1} \right),$$

Step 6. Sampling the random effects \mathbf{u} . let $\boldsymbol{\theta}_i^* = \boldsymbol{\theta}_i - \gamma_{0,g} - \mathbf{t}_i \gamma_{1,g} - \mathbf{w}_i \gamma_2$. Let $\mathbf{z}_i = (\mathbf{1}, \mathbf{t}_i, \mathbf{w}_i)$. Then, the posterior distribution is given by

$$\mathbf{u}_i | \mathbf{z}_i, \boldsymbol{\theta}, \sigma, \boldsymbol{\tau}_g, \boldsymbol{\gamma}_g \sim \text{MVN} \left(\frac{\mathbf{z}_i^t \sigma^{-2} \boldsymbol{\theta}_i^*}{\sigma^2 (\mathbf{z}_i^t \mathbf{z}_i)^{-1} + \boldsymbol{\tau}_g^{-1}}, (\sigma^2 (\mathbf{z}_i^t \mathbf{z}_i)^{-1} + \boldsymbol{\tau}_g^{-1})^{-1} \right).$$

Step 7. Sampling the parameters of the survival model. For the regression parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}$, the functional form of the likelihood is determined by the choice of the density function (exponential, Weibull or lognormal). With the normal prior, a Metropolis–Hastings step was implemented with a normal proposal density. The proposal was evaluated against the previous draw similar as outlined in Step 3. The same applies for $\boldsymbol{\Lambda}$. For the variance parameters of either the Weibull or lognormal model, the proposal was drawn from a Gamma distribution and the updating step was again a Metropolis–Hastings step.

Step 8. The variance parameter σ^2 was sampled from an inverse-gamma distribution, and $\boldsymbol{\tau}_g$ from an inverse-Wishart distribution using conjugated priors.

To get good acceptance rates for the several Metropolis–Hastings steps in the algorithm, the acceptance rate over a period of 100 iterations were computed. Then, the acceptance rate was evaluated. If the acceptance rate fell below 0.20 the variance of the proposal distribution was adjusted to a smaller value, if the acceptance rate became above 0.50, the variance of the proposal was set to a larger value. This adaptive Metropolis–Hastings procedure was only used in the burn-in period.

References

- Larsen K. The Cox proportional hazards model with a continuous latent variable measured by multiple binary indicators. *Biometrics* 2005; **61**:1049–1055.
- Wang C, Douglas J, Anderson S. Item response models for joint analysis of quality of life and survival. *Statistics in Medicine* 2002; **21**:129–142. DOI: 10.1002/sim.989.
- Jiang Y, Hesser JE. Using item response theory to analyze the relationship between health-related quality of life and health risk factors. *Preventing Chronic Disease* 2009; **6**(1):1–10.
- Schulz-Larsen K, Kreiner S, Lomholt RK. Mini-mental status examination: mixed rasch model item analysis derived two different cognitive dimensions of the mmse. *Journal of Clinical Epidemiology* 2007; **60**:268–279. DOI: 10.1016/j.jclinepi.2006.06.007.
- Meredith W, Tisak J. Latent curve analysis. *Psychometrika* 1990; **55**:107–122.
- Duncan TE, Duncan SC, Strycker LA, Li F, Alpert A. *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Applications*. Erlbaum: Mahwah, NJ, 1999.
- van der Linden WJ, Hambleton RK (eds). *Handbook of Modern Item Response Theory*. Springer: New York, 1997.
- Muraki E. A generalized partial credit model: application of an EM algorithm. *Applied Psychological Measurement* 1992; **16**:159–176. DOI: 10.1177/014662169201600206.
- Muraki E. Information functions of the generalized partial credit model. *Applied Psychological Measurement* 1993; **17**:351–363. DOI: 10.1177/014662169301700403.
- Fox J-P, Glas CAW. Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika* 2001; **66**:271–288.
- Goodman LA. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* 1974; **61**:215–231. DOI: 10.1093/biomet/61.2.215.
- Rost J. Rasch models in latent classes: an integration of two approaches to item analysis. *Applied Psychological Measurement* 1990; **14**:271–282. DOI: 10.1177/014662169001400305.
- Nagin DS. Analyzing developmental trajectories: a semiparametric, group based approach. *Psychological Methods* 1999; **4**:139–157.
- Nagin DS, Tremblay RE. Analyzing developmental trajectories of distinct but related behaviors: a group-based method. *Psychological Methods* 2001; **6**:18–34.
- Fox J-P. *Bayesian Item Response Modeling*. Springer: New York, 2010.
- Vermunt J. Latent class and finite mixture models for multilevel datasets. *Statistical Methods in Medical Research* 2008; **17**:33–51. DOI: 10.1177/0962280207081238.
- Asparouhov T, Masyn K, Muthén B. Continuous time survival in latent variable models. *Proceedings of the American Statistical Association, Biometrics Section*. American Statistical Association: Alexandria, VA, 2006; 180–187.
- Xu J, Zeger SL. Joint analysis of longitudinal data comprising repeated measures and time to events. *Journal of the Royal Statistical Society, Series C* 2001; **50**:375–387.
- Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**:330–339.
- Box-Steffenmeier JM, Zorn CJW. Duration models and proportional hazards in political science. *American Journal of Political Science* 2001; **45**:972–988.
- Tsiatis A, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* 2004; **14**:809–834.

22. Prentice RL. Covariate measurement errors and parametric estimation in a failure time regression model. *Biometrika* 1982; **69**:331–342. DOI: 10.1093/biomet/69.2.331.
23. Tombaugh TN. The mini-mental state examination: a comprehensive review. *The Journal of the American Geriatrics Society* 1992; **40**:922–935.
24. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**:963–74.
25. Fox J-P, Glas CAW. Bayesian modeling of measurement error in predictor variables using item response theory. *Psychometrika* 2003; **68**:169–191.
26. Ibrahim J, Chen MH, Sinha D. *Bayesian Survival Analysis*. Springer: New York, 2001.
27. Larsen K. Joint analysis of time-to-event and multiple binary indicators of latent classes. *Biometrics* 2004; **60**:85–92.
28. Hogan JW, Laird NM. Mixture models for the joint distribution of repeated measures and event time. *Statistics in Medicine* 1997; **16**:239–257. DOI: 10.1002/(SICI)1097-0258(19970215).
29. Geweke J. Interpretation and inference in mixture models: simple MCMC works. *Computational Statistics and Data Analysis* 2007; **51**:3529–3550. DOI: 10.1016/j.csda.2006.11.026.
30. Smith BJ. boa: an R package for MCMC output convergence assessment and posterior inference. *Journal of Statistical Software* 2007; **21**:1–37.
31. Gelman A, Meng XL, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* 1996; **6**:733–807.
32. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis* (2nd edn). Chapman & Hall: New York, 2004.
33. Sinharay S. Assessing fit of unidimensional item response theory models using a Bayesian approach. *Journal of Educational Measurement* 2005; **42**:375–394. DOI: 10.1111/j.1745-3984.2005.00021.x.
34. Hanson TE, Branscum AJ, Johnson WO. Predictive comparison of joint longitudinal-survival modeling: a case study illustrating competing approaches. *Lifetime Data Analysis* 2011; **17**:3–28. DOI: 10.1007/s10985-010-9162-0.
35. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B* 2002; **64**:583–639. DOI: 10.1111/1467-9868.00353.
36. Sinharay S, Johnson MS, Stern HS. Posterior predictive assessment of item response theory models. *Applied Psychological Measurement* 2006; **30**:298–321. DOI: 10.1177/0146621605285517.
37. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data* (2nd edn). Wiley: Hoboken, NJ, 2002.
38. Neale R, Brayne C, Johnson A. Cognition and survival: an exploration in a large multicentre study of the population aged 65 years and over. *International Journal of Epidemiology* 2001; **30**(6):1383–1388. DOI: 10.1093/ije/30.6.1383.