Web-based Supporting Materials for "Functional Joint Model for Longitudinal and Time-to-Event Data: An Application to Alzheimer's Disease"

1 Preliminary analysis of the ADNI Study

In this preliminary analysis, we simultaneously model the longitudinal ADAS-Cog 11 score and time to AD diagnosis in a joint modeling framework, with the baseline hippocampal volume (bHV) as a scalar covariate, because degree of atrophy within the medial temporal lobe structures, especially within the hippocampus, was reported to be sensitive to the disease progression.

Model 1 (refer to as $JM1$) includes baseline age ($bAge$), gender ($gender$), years of education (Edu), and presence of the apolipoprotein E allele ($APOE$–$\varepsilon 4$) as scalar covariates given their potential effects on disease progression in AD [1–3].

$$\begin{aligned} ADAS\text{–}Cog_i(t_{ij}) &= m_i(t_{ij}) + \varepsilon_{ij} \\ m_i(t_{ij}) &= \beta_0 + \beta_1 t_{ij} + \beta_2 bAge_i + u_i \\ h(t) &= h_0(t)\exp\{\gamma_1 gender_i + \gamma_2 bAge_i + \gamma_3 Edu_i + \gamma_4 APOE\text{–}\varepsilon 4 + \alpha m_i(t)\}, \end{aligned}$$

where random effects $u_i \sim N(0, \sigma_u^2)$. Model 2 (referred to as $JM2$) and Model 3 (referred to as $JM3$) are similar to model $JM1$, but include variable bHV in the longitudinal submodel and survival submodel, respectively. Model 4 (referred to as $JM4$) is similar to model $JM1$, but includes variable bHV in both the longitudinal and survival submodels.

	$JM1$	$JM2$	$JM3$	$JM4$
AIC	10525	10472	10514	10454

Table 1: Akaike information criterion (AIC) from four joint models for the ADNI study.

Table 1 displays the values of AIC from these four models. All three joint models with baseline hippocampal volume (bHV) as a covariate (models $JM2$, $JM3$, and $JM4$) have smaller AIC values than model $JM1$, indicating that incorporating hippocampal volume information, in addition to clinical and genetic characteristics, can improve the joint model fit. Model $JM4$ yields the smallest AIC value of 10454, suggesting that the baseline hippocampal volume is associated with both ADAS-Cog 11 score and time to AD diagnosis. The results from model $JM4$ are presented in Table 2.

	Parameters	MLE	SE	p
For longitudinal outcome				
ADAS-Cog 11	Time (Years)	0.425	0.047	<0.001
	$bAge$	−0.389	0.254	0.125
	bHV (mm^3)	−1.878	0.221	<0.001
For survival process				
MCI to AD	Female	−0.161	0.167	0.331
	$bAge$	−0.181	0.0867	0.037
	Edu (years)	−0.004	0.026	0.866
	$APOE$-ε	0.397	0.167	0.018
	bHV (mm^3)	−3.559	0.928	<0.001
	α	0.108	0.019	<0.001

Table 2: ADNI data analysis results from model $JM4$ with baseline hippocampal volume (bHV) as covariates in both longitudinal and survival submodels. Parameter α is the association parameter which measures the strength of the association between the trajectory of ADAS-Cog 11 and time to AD diagnosis.

2 Hippocampus Image Processing

For image processing, we adopt a surface fluid registration package [4] which has been used in various studies [5–8]. The original imaging data, 3D brain MRI scans, are downloaded from `www.loni.ucla.edu/ADNI`. The hippocampal surfaces are automatically segmented using FMRIB's Integrated Registration and Segmentation Tool) (FIRST) [9], an integrated surface analysis tool developed as part of the FSL library [10]. After the segmentation, the cortical surface of left and right hippocampi are extracted. The surface is the outer layer of the brain region and has an inherent 2D structure. Then the surface is modeled as a mesh of triangles for each side of the hippocampi. Each triangle is know as a face. The place where the corners of the triangles meet is called a vertex. The coordinates (i.e., the X, Y, and Z) at each vertex are determined from the MRI during the extraction process. In order to do the registration, the surfaces are first conformally mapped to a rectangle plane using holomorphic 1-forms. This process is similar to unfolding a paper bag with edges being sealed together. A feature images of the surfaces is computed from this conformal representation, and are registered to a chosen template image via inverse consistent surface fluid registration. Using conformal mapping, the surface registration problem is essentially converted into an image registration problem. Because of the one-to-one correspondence of points on the feature image and the vertices on the surface mesh, the registered feature image can be recovered in the original hippocampal surfaces. Detailed image processing and registration procedure can be found in Shi *et al* [4].

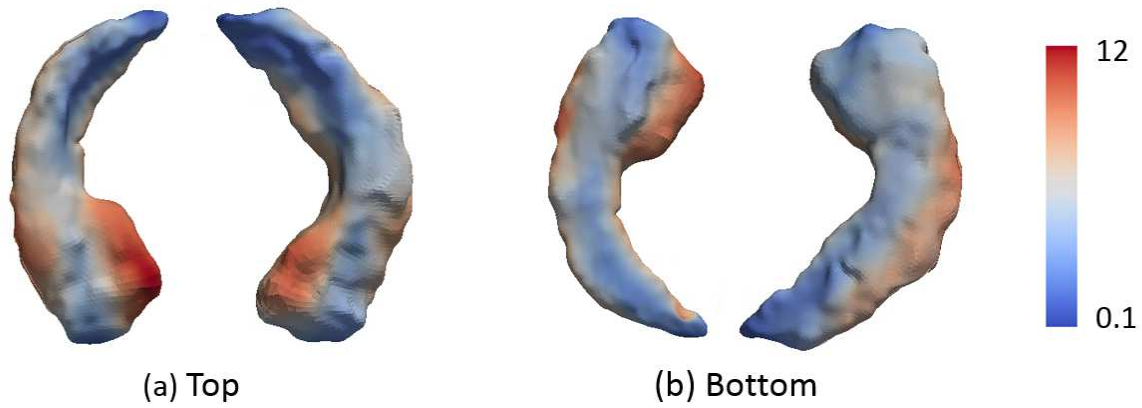

Figure 1: The top and bottom views of one MCI patient's hippocampal radial distance denoted by colors.

Each registered surface corresponding to the either left or right hippocampal surfaces contains 15,000 vertices. Knowledge of the coordinates of the vertices allows us to compute several surface statistics, including radial distance, multivariate tensor-based morphometry (mTBM), determinant of the Jacobian matrix, and two eigenvalues of the Jacobian matrix. Specifically, we adopt radial distance as our target measurement. Figure 1 displays the hippocampus surface morphology of one patient with radial distance being coded in colors. We align the points on the two-dimensional feature image to a one-dimensional domain (denoted by $\boldsymbol{S}$) in a particular order, where the order is preserved across all vertices. The corresponding radial distances of vertices are thus vectorized to form a one-dimensional functional predictor denoted by $g_i^{(x)}(s)$. The coefficient function $B^{(x)}(s)$ is defined on the same domain as $g_i^{(x)}(s)$. The one-to-one correspondence between the points on domain $\boldsymbol{S}$ and the vertices on surfaces allows us to map $g_i^{(x)}(s)$ and $B^{(x)}(s)$ to the hippocampal surface easily. Figure 2 displays the data processing procedure along with an illustration of a functional joint model.

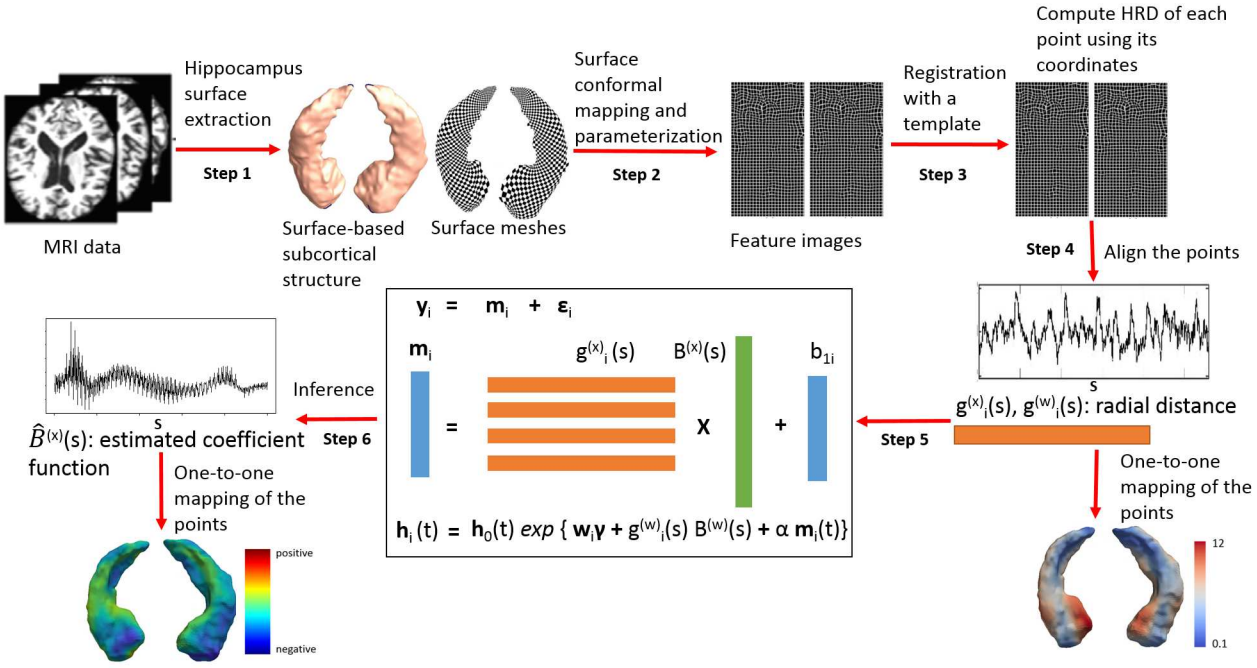

Figure 2: Hippocampus image processing procedure along with an illustration of a functional joint model.

3 Implementing the EM Algorithm

The estimation of the joint likelihood function $L(\boldsymbol{\theta}) = \prod_{i=1}^{I} p(\boldsymbol{y}_i, T_i, \delta_i|\boldsymbol{\theta})$ can be achieved by using Expectation-Maximization (EM) algorithms, where in the E-step the unknown random effects are treated as missing values. The log-likelihood function for the observed data is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{I} \big\{ \log p(\boldsymbol{y}_i|\boldsymbol{\theta}, \boldsymbol{u}_i) + \log p(T_i, \delta_i, |\boldsymbol{\theta}, \boldsymbol{u}_i) + \log p(\boldsymbol{u}_i; \boldsymbol{\theta}) \big\}.$$

To simplify the notation in models (3) and (4), we let $\boldsymbol{x}_i^*(t_{ij}) = \big[1, \boldsymbol{x}_{ij}^\top, (\boldsymbol{\xi}_i^{(x)})^\top\big]^\top$ as the covariates vector recorded at time t_{ij} from the study onset, $\boldsymbol{\beta}^* = \big[\beta_0', \boldsymbol{\beta}^\top, (\boldsymbol{B}^{(x)})^\top\big]^\top$, $\boldsymbol{z}_i(t_{ij}) = \boldsymbol{z}_{ij}$, $\boldsymbol{w}_i^* = \big[\boldsymbol{w}_i^\top, (\boldsymbol{\xi}_i^{(w)})^\top\big]^\top$, and $\boldsymbol{\gamma}^* = \big[\boldsymbol{\gamma}^\top, (\boldsymbol{B}^{(w)})^\top\big]^\top$. Then models (3) and (4) can be expressed as

$$\begin{aligned} y_i(t) &= m_i(t) + \varepsilon_{ij}, \text{ where } m_i(t) = \boldsymbol{x}_i^*(t)^\top \boldsymbol{\beta}^* + \boldsymbol{z}_i(t)^\top \boldsymbol{u}_i, \\ \text{and } h(t) &= h_0^*(t) \exp\{\boldsymbol{w}_i^{*\top} \boldsymbol{\gamma}^* + \alpha m_i(t)\}. \end{aligned}$$

In the E-step, the expected complete log-likelihood function given the conditional distribution of random effect is

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) = \sum_{i=1}^{I}\int \big\{\log p(\boldsymbol{y}_i|\boldsymbol{\theta},\boldsymbol{u}_i)+\log p(T_i,\delta_i,|\boldsymbol{\theta},\boldsymbol{u}_i)+\log p(\boldsymbol{u}_i;\boldsymbol{\theta})\big\}p(\boldsymbol{u}_i|\boldsymbol{y}_i,T_i,\delta_i;\boldsymbol{\theta}^{(m)})d\boldsymbol{u}_i,$$

where $\boldsymbol{\theta}^{(m)}$ denotes the parameters at the current iteration. In the M-step, the residual variance and the covariance matrix of the random effects $\boldsymbol{u}_i$ in the longitudinal submodel can be updated using closed-form expressions,

$$\begin{aligned}\widehat{\sigma}^2_{\epsilon} &= \frac{1}{\sum_{i=1}^{I}J_i}\sum_{i=1}^{I}\big(\boldsymbol{y}_i-\boldsymbol{X}^*_i\boldsymbol{\beta}^*\big)^{\top}\big(\boldsymbol{y}_i-\boldsymbol{X}^*_i\boldsymbol{\beta}^*-2\boldsymbol{Z}_iE(\boldsymbol{u}_i|\boldsymbol{y}_i,T_i,\delta_i;\boldsymbol{\theta}^{(m)})\big)\\ &\quad +tr\big(\boldsymbol{Z}_i^{\top}\boldsymbol{Z}_iVar(\boldsymbol{u}_i|\boldsymbol{y}_i,T_i,\delta_i;\boldsymbol{\theta}^{(m)})\big)+E(\boldsymbol{u}_i|\boldsymbol{y}_i,T_i,\delta_i;\boldsymbol{\theta}^{(m)})^{\top}\boldsymbol{Z}_i^{\top}\boldsymbol{Z}_iE(\boldsymbol{u}_i|\boldsymbol{y}_i,T_i,\delta_i;\boldsymbol{\theta}^{(m)}),\\ \widehat{\boldsymbol{\Sigma}}_u &= \frac{1}{\sum_{i=1}^{I}J_i}\sum_{i=1}^{I}Var(\boldsymbol{u}_i|\boldsymbol{y}_i,T_i,\delta_i;\boldsymbol{\theta}^{(m)})+E(\boldsymbol{u}_i|\boldsymbol{y}_i,T_i,\delta_i;\boldsymbol{\theta}^{(m)})E(\boldsymbol{u}_i|\boldsymbol{y}_i,T_i,\delta_i;\boldsymbol{\theta}^{(m)})^{\top},\end{aligned}$$

where $\boldsymbol{X}^*_i=\big[\boldsymbol{x}^{*\top}_i(t_{i1}),\cdots,\boldsymbol{x}^{*\top}_i(t_{iJ_i})\big]^{\top}$, $\boldsymbol{Z}_i=\big[\boldsymbol{z}_{i1},\cdots,\boldsymbol{z}_{iJ_i}\big]^{\top}$, tr is the trace function of a matrix, E denotes the expectation function, and Var represents the variance function.

There are no closed-form solutions for the fixed effects $\boldsymbol{\beta}^*$ and the parameters of the survival submodel. The one-step Newton-Raphson algorithm can be implemented to update the parameters.

$$\widehat{\boldsymbol{\beta}}^{*(m+1)}=\widehat{\boldsymbol{\beta}}^{*(m)}-(\partial S(\widehat{\boldsymbol{\beta}}^{*(m)})/\partial\boldsymbol{\beta}^{\top})^{-1}S(\widehat{\boldsymbol{\beta}}^{*(m)}),$$

where $\widehat{\boldsymbol{\beta}}^{*(m)}$ denotes the value of $\boldsymbol{\beta}^*$ at the current iteration, and the corresponding score vector is

$$\begin{aligned}S(\boldsymbol{\beta}^*) &= \sum_{i=1}^{I}\frac{1}{\sigma^2_{\epsilon}}\boldsymbol{X}^{*\top}_i\big(\boldsymbol{y}_i-\boldsymbol{X}^*_i\boldsymbol{\beta}^*-\boldsymbol{Z}_iE(\boldsymbol{u}_i|\boldsymbol{y}_i,T_i,\delta_i;\boldsymbol{\theta}^{(m)})\big)+\delta_i\alpha\boldsymbol{x}^*_i(T_i)\\ &\quad -\exp(\boldsymbol{w}^{*\top}_i\boldsymbol{\gamma}^*)\int\int_0^{T_i}h^*_0(t)\alpha\boldsymbol{x}^*_i(t)\exp\big(\boldsymbol{x}^*_i(t)^{\top}\boldsymbol{\beta}^*+\boldsymbol{z}_i(t)^{\top}\boldsymbol{u}_i\big)\\ &\quad \times p(\boldsymbol{u}_i|\boldsymbol{y}_i,T_i,\delta_i;\boldsymbol{\theta}^{(m)})\big)dtd\boldsymbol{u}_i.\end{aligned}$$

The $\partial S(\widehat{\boldsymbol{\beta}}^{*(m)})/\partial\boldsymbol{\beta}^{\top}$ denotes the corresponding blocks of the Hessian matrix, and is approximated using forward difference methods. The integral with respect to time and the integral with respect to random effects are evaluated via a pseudo-adoptive Gaussian-Hermit quadrature rule [11].

Parameters in the survival submodel are updated in a similar fashion. The score vectors corresponding to $\boldsymbol{\gamma}^*$, α, $\boldsymbol{\theta}_{h_0^*}$ are

$$\begin{aligned} S(\boldsymbol{\gamma}^*) &= \sum_{i=1}^{I} \boldsymbol{w}_i^* \big[\delta_i - \exp(\boldsymbol{w}_i^{*\top}\boldsymbol{\gamma}^*) \int\int_0^{T_i} h_0^*(t) \exp\big(\alpha\{\boldsymbol{x}_i^*(t)^\top \boldsymbol{\beta}^* + \boldsymbol{z}_i(t)^\top \boldsymbol{u}_i\}\big) \\ &\quad \cdot p(\boldsymbol{u}_i|\boldsymbol{y}_i, T_i, \delta_i; \boldsymbol{\theta}^{(m)})\big) dt d\boldsymbol{u}_i\big], \end{aligned}$$

$$\begin{aligned} S(\alpha) &= \sum_{i=1}^{I} \delta_i \{\boldsymbol{x}_i^*(T_i)^\top \boldsymbol{\beta}^* + \boldsymbol{z}_i(T_i)^\top \boldsymbol{u}_i\} \\ &\quad - \exp(\boldsymbol{w}_i^{*\top}\boldsymbol{\gamma}^*) \int\int_0^{T_i} h_0^*(t) \exp\big(\alpha\{\boldsymbol{x}_i^*(t)^\top \boldsymbol{\beta}^* + \boldsymbol{z}_i(t)^\top \boldsymbol{u}_i\}\big) p(\boldsymbol{u}_i|\boldsymbol{y}_i, T_i, \delta_i; \boldsymbol{\theta}^{(m)}) dt d\boldsymbol{u}_i, \end{aligned}$$

$$\begin{aligned} S(\boldsymbol{\theta}_{h_0^*}) &= \sum_{i=1}^{I} \delta_i \frac{\partial h_0^*(T_i; \boldsymbol{\theta}_{h_0^*})}{\partial \boldsymbol{\theta}_{h_0^*}^\top} \\ &\quad - \exp(\boldsymbol{w}_i^{*\top}\boldsymbol{\gamma}^*) \int\int_0^{T_i} \frac{\partial h_0^*(t; \boldsymbol{\theta}_{h_0^*})}{\partial \boldsymbol{\theta}_{h_0^*}^\top} \exp\big(\alpha\{\boldsymbol{x}_i^*(t)^\top \boldsymbol{\beta}^* + \boldsymbol{z}_i(t)^\top \boldsymbol{u}_i\}\big) p(\boldsymbol{u}_i|\boldsymbol{y}_i, T_i, \delta_i; \boldsymbol{\theta}^{(m)}) dt d\boldsymbol{u}_i. \end{aligned}$$

The E-step and M-step iterate until a pre-specified convergence criterion is met, and the maximum likelihood estimator (MLE) $\widehat{\boldsymbol{\theta}}$ is obtained.

By employing forward difference approximation, the Hessian matrix $\partial S(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^\top$ can be calculated using the functions that compute the score vector. The estimated observed information matrix is achieved by $I(\widehat{\boldsymbol{\theta}}) = -\partial S(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^\top|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$, and subsequently variance-covariance matrix for the MLE via $\widehat{Var}(\widehat{\boldsymbol{\theta}}) = I(\widehat{\boldsymbol{\theta}})^{-1}$.

4 Application to the ADNI Study

Parameters	MLE	SE	p
$B_1^{(x)}$	1.778	0.424	<0.001
$B_2^{(x)}$	−0.528	0.546	0.333
$B_3^{(x)}$	0.122	0.600	0.839
$B_4^{(x)}$	1.796	0.752	0.017
$B_5^{(x)}$	1.052	0.857	0.220
$B_6^{(x)}$	−1.724	0.994	0.083
$B_7^{(x)}$	−1.406	0.954	0.140
$B_8^{(x)}$	1.629	1.234	0.187
$B_9^{(x)}$	3.228	1.086	0.003
$B_{10}^{(x)}$	−6.743	1.300	<0.001
$B_{11}^{(x)}$	−4.043	1.506	0.007
$B_{12}^{(x)}$	−1.955	1.343	0.146
$B_{13}^{(x)}$	1.823	1.423	0.200
$B_{14}^{(x)}$	−1.451	1.574	0.357
$B_{15}^{(x)}$	−3.518	1.475	0.017
$B_{16}^{(x)}$	−5.205	1.808	0.004
$B_{17}^{(x)}$	−0.165	1.569	0.916
$B_{18}^{(x)}$	9.129	1.943	<0.001
$B_{19}^{(x)}$	3.757	2.034	0.065
$B_{20}^{(x)}$	3.548	1.830	0.053

Table 3: The estimated coefficients for the 20 FPC scores derived from hippocampal radial distance (HRD). Model: the proposed functional joint model $FJM1$ with HRD as a functional predictor in the longitudinal submodel.

References

[1] Risacher SL, Saykin AJ, Wes JD, Shen L, Firpi HA, McDonald BC. Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. Current Alzheimer Research. 2009;6(4):347–361.

[2] Cui Y, Liu B, Luo S, Zhen X, Fan M, Liu T, et al. Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. PLOS ONE. 2011;6(7):e21896.

[3] Li S, Okonkwo O, Albert M, Wang MC. Variation in variables that predict progression from MCI to AD dementia over duration of follow-up. American Journal of Alzheimer's Disease. 2013;2(1):12–28.

[4] Shi J, Thompson PM, Gutman B, Wang Y, Initiative ADN, et al. Surface fluid registration of conformal representation: Application to detect disease burden and genetic influence on hippocampus. NeuroImage. 2013;78:111–134.

[5] Colom R, Stein JL, Rajagopalan P, Martínez K, Hermel D, Wang Y, et al. Hippocampal structure and human cognition: Key role of spatial processing and evidence supporting the efficiency hypothesis in females. Intelligence. 2013;41(2):129–140.

[6] Luders E, Thompson PM, Kurth F, Hong JY, Phillips OR, Wang Y, et al. Global and regional alterations of hippocampal anatomy in long-term meditation practitioners. Human Brain Mapping. 2013;34(12):3369–3375.

[7] Monje M, Thomason ME, Rigolo L, Wang Y, Waber DP, Sallan SE, et al. Functional and structural differences in the hippocampus associated with memory deficits in adult survivors of acute lymphoblastic leukemia. Pediatric Blood & Cancer. 2013;60(2):293–300.

[8] Shi J, Wang Y, Ceschin R, An X, Lao Y, Vanderbilt D, et al. A multivariate surface-based analysis of the putamen in premature newborns: regional differences within the ventral striatum. PLOS ONE. 2013;8(7):e66736.

[9] Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. NeuroImage. 2011;56(3):907–922.

[10] Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. NeuroImage. 2012;62(2):782–790.

[11] Rizopoulos D. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. Computational Statistics & Data Analysis. 2012;56(3):491–501.