

Improved Dynamic Predictions from Joint Models of Longitudinal and Survival Data with Time-Varying Effects using P-splines

E.R. Andrinopoulou^{1,*}, P.H.C. Eilers^{1,***}, J.J.M. Takkenberg^{2,****}, and D. Rizopoulos^{1,**}

¹Department of Biostatistics, Erasmus MC, Rotterdam, The Netherlands

²Department of Cardiothoracic Surgery, Erasmus MC, Rotterdam, The Netherlands

**email*: e.andrinopoulou@erasmusmc.nl

***email*: p.eilers@erasmusmc.nl

****email*: j.j.m.takkenberg@erasmusmc.nl

*****email*: d.rizopoulos@erasmusmc.nl

SUMMARY: In the field of cardio-thoracic surgery, valve function is monitored over time after surgery. The motivation for our research comes from a study which includes patients who received a human tissue valve in the aortic position. These patients are followed prospectively over time by standardized echocardiographic assessment of valve function. Loss of follow-up could be caused by valve intervention or the death of the patient. One of the main characteristics of the human valve is that its durability is limited. Therefore, it is of interest to obtain a prognostic model in order for the physicians to scan trends in valve function over time and plan their next intervention, accounting for the characteristics of the data.

Several authors have focused on deriving predictions under the standard joint modeling of longitudinal and survival data framework that assumes a constant effect for the coefficient that links the longitudinal and survival outcomes. However, in our case this may be a restrictive assumption. Since the valve degenerates, the association between the biomarker with survival may change over time.

To improve dynamic predictions we propose a Bayesian joint model that allows a time-varying coefficient to link the longitudinal and the survival processes, using P-splines. We evaluate the performance of the model in terms of discrimination and calibration, while accounting for censoring.

KEY WORDS: calibration; discrimination; joint model; longitudinal outcome; P-splines; survival outcome.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

In the field of cardio-thoracic surgery, valve function is monitored periodically over time after heart valve surgery. Aortic gradient (AG) (mmHg) is one of the continuous echocardiographic markers that measures valve (dys)function, where high values indicate a worsening of the patient's condition. Specifically, it measures aortic stenosis which occurs when the opening of the aortic valve located between the left ventricle of the heart and the aorta is narrowed. During the follow-up period after surgery, patients may require an intervention or may die. The motivation of this research comes from a study, conducted in the Erasmus University Medical Center, which includes all patients who received a human tissue valve allograft in the aortic position in the Department of Cardio-Thoracic Surgery (Bekkers et al., 2011). In total 296 patients who survived aortic valve or root replacement with an allograft valve were followed over time. Specifically, echocardiographic examinations were scheduled at six months, one year postoperatively and biennially thereafter. During follow-up, 161 (54%) patients either died or required a reoperation on the same valve. A total of 1669 echocardiographic measurements of AG were performed. The median number of visits is six and the median years of follow up is 9.3. One of the characteristics of human valves is that their durability is limited. Hence, it is important for the physicians to have a prognostic tool in order to carefully monitor trends in valve function over time and plan a future re-intervention.

Joint modeling of longitudinal and survival data is a popular framework to analyze data including repeated measurements and time-to-event outcomes appropriately (Tsiatis and Davidian, 2004; Rizopoulos, 2012; Andrinopoulou et al., 2012; Rizopoulos et al., 2015). The idea behind these models is that the longitudinal and the survival processes share common random effects, inducing correlation between the two processes. Specifically, we construct a mixed-effects model to describe the evolution over time for the longitudinal outcome, and use these estimated evolutions as a time-dependent covariate in a survival model. Several authors

have focused on deriving predictions under joint modeling of longitudinal and survival data framework (Taylor, Yu, and Sandler, 2005; Garre et al., 2008; Yu, Taylor, and Sandler, 2008; Proust-Lima and Taylor, 2009; Rizopoulos, 2011; Andrinopoulou et al., 2015a). To improve the fit and the predictive accuracy of joint models, previous work allowed the inclusion of multiple longitudinal outcomes and investigated the selection of the optional functional form (Rizopoulos, 2011; Rizopoulos et al., 2014; Andrinopoulou et al., 2015b; Andrinopoulou and Rizopoulos, 2016). A common feature of all aforementioned models and previous work published on the motivating data is that the parameters that measure the strength of the association between the longitudinal and survival outcome were assumed to be constant in time. However, the heart valve degenerates, therefore it is natural and biologically more desirable to assume that the effect of AG is changing over time. To investigate that, we performed a preliminary analysis assuming a time-dependent Cox model for the composite event death/reoperation including as predictors the square root of AG (SAG), a transformation that was explored in previous work (Andrinopoulou et al., 2014), and gender. Figure 1 shows the smoothed scatterplot of the Schoenfeld residuals from this Cox model versus time. These residuals are typically used to investigate the proportional hazards assumption, and for the SAG they show an increasing trend indicating violation of a constant-effect assumption.

Even though the consideration of time-varying coefficients has been extensively studied in the general context of survival analysis using polynomials and B-splines (Nan et al., 2005; Perperoglou, 2014), relatively little work has been done on joint models with time-varying coefficients (Song and Wang, 2008). To our knowledge, no work has been done to evaluate whether such time-varying coefficients may improve the accuracy of individualized predictions within the framework of joint models. The idea behind the time-varying coefficient models is to include interactions of the covariates with an appropriate pre-defined time

function. To enhance the predictive performance of the joint model, we assume a time-varying effect of SAG using P-splines. Specifically, it is approximated by a polynomial spline written in terms of a linear combination of B-spline basis functions (Eilers and Marx, 1996; Lang and Brezger, 2004; Eilers, Marx, and Durbán, 2015). To overcome the problem of the large number of parameters and to stabilize the predictions, a penalty is applied to the coefficients. To facilitate flexible modeling of the survival outcome, we use P-splines also for the logarithm of the baseline hazard. To evaluate the derived predictions we present extensions of classic measures of predictive ability, such as discrimination and calibration, in the time-dependent setting while accounting for censoring.

The rest of the paper is organized as follows. Section 2 describes the formulation of the joint model. Section 3 presents the Bayesian estimation. Section 4 presents measures to assess the predictive performance of the model. Section 5 shows the results for the cardio data analysis, while Section 6 contains simulation studies. Finally, in Section 7 we close with a discussion.

[Figure 1 about here.]

2. Joint Model Definition

We let T_i^* denote the true failure time for the i -th individual ($i = 1, \dots, n$), and C_i the censoring time. Moreover, $T_i = \min(T_i^*, C_i)$ denotes the observed failure time and $\delta_i = \{0, 1\}$ is the event indicator where zero indicates censoring. We let \mathbf{y}_i denote the longitudinal response obtained at different time points $t_{ij} > 0$, ($j = 1, \dots, n_i$). To describe the subject-specific evolution over time of the continuous longitudinal outcome SAG, we use a mixed-effects model. In particular, we postulate

$$y_i(t) = \eta_i(t) + \epsilon_i = \mathbf{x}_i^\top(t)\boldsymbol{\beta} + \mathbf{z}_i^\top(t)\mathbf{b}_i + \epsilon_i(t),$$

where $\mathbf{x}_i(t)$ denotes the design vector for the fixed effects regression coefficients $\boldsymbol{\beta}$ and $\mathbf{z}_i(t)$ the design vector for the random effects \mathbf{b}_i . Moreover, $\epsilon_i(t) \sim N(0, \sigma)$. For the corresponding random effects, we assume a multivariate normal distribution, namely

$$\mathbf{b}_i \sim N_b(\mathbf{0}, \boldsymbol{\Sigma}_b^2).$$

We postulate a varying-coefficient joint model (VCJM) for the relationship between the survival and the longitudinal outcome. Specifically, we have

$$h_i(t, \boldsymbol{\theta}_s) = h_0(t) \exp[\boldsymbol{\gamma}^\top \mathbf{w}_i + f\{\lambda(t), \mathcal{H}_i(t)\}],$$

where $\boldsymbol{\theta}_s$ is the parameter vector for the survival outcomes, \mathbf{w}_i is a vector of baseline covariates with a corresponding vector of regression coefficients $\boldsymbol{\gamma}$, $\boldsymbol{\gamma}_{h_0}$ is the vector of the baseline hazard coefficient and $\mathcal{H}_i(t) = \{\eta_i(\zeta), 0 \leq \zeta < t\}$ denotes the history of the true unobserved longitudinal process up to time point t . The function $f\{\lambda(t), \mathcal{H}_i(t)\}$ specifies which features of the longitudinal submodel are included in the relative risk model (Rizopoulos and Ghosh, 2011; Rizopoulos, 2012). Previous work suggested that not only the value of SAG but also other characteristics of the biomarker may have an influence on the event of interest (Andrinopoulou et al., 2015b). Several specifications of $f(\cdot)$ have been proposed in the literature (Rizopoulos and Ghosh, 2011). Some examples are the following:

$$\begin{aligned} f\{\lambda(t), \mathcal{H}_i(t)\} &= \lambda(t)\eta_i(t), \\ f\{\lambda(t), \mathcal{H}_i(t)\} &= \lambda_1(t)\eta_i(t) + \lambda_2(t)\frac{d\eta_i(t)}{dt}, \\ f\{\lambda(t), \mathcal{H}_i(t)\} &= \lambda(t) \int_0^t \eta_i(s)ds. \end{aligned}$$

Specifically, $f(\cdot)$ postulates that the hazard of the event is associated with the underlying value of the longitudinal outcome at a specific time point t , the value and the slope of the longitudinal outcome at t , or the accumulated longitudinal process up to time t .

In the standard constant-coefficient joint model (CCJM), $\lambda(t)$ is assumed to be constant over time. For the VCJM, we model $\lambda(t)$ flexibly assuming a smooth function. Several approaches have been proposed for modelling and estimating smooth functions such as B-

splines. A problem that arises in such methods is the selection of the number and position of the knots which has been a subject of much research. In this manuscript, we adopt the P-splines approach for $\lambda(t)$. The basic idea of P-splines is to use a (relatively) high number of equally spaced knots. To obtain sufficient smoothness of the fitted curves and to avoid overfitting, a roughness penalty based on differences of adjacent B-spline coefficients is applied (Eilers and Marx, 1996). In particular, we take

$$\lambda(t) = \sum_{\ell=1}^L \alpha_{\ell} B_{\ell}(t),$$

where α_{ℓ} is a set of parameters that capture the strength of association between the longitudinal and survival outcomes and $B_{\ell}(t)$ denotes the ℓ -th basis function of a B-spline. To further facilitate improving the derived predictions from the joint model, we model the baseline hazard with the same P-splines approach. Specifically,

$$\log\{h_0(t)\} = \sum_{u=1}^U \gamma_{h_0,u} B_u(t),$$

where $\gamma_{h_0,u}$ are the coefficients of the baseline hazard and $B_u(t)$ denotes the u -th basis function of a B-spline.

3. Bayesian Estimation

We employ a Bayesian approach where inference is based on the posterior of the model. In particular, we use Markov chain Monte Carlo (MCMC) methods to estimate the parameters of the VCJM. The likelihood of the model is derived under the assumption that the longitudinal and survival processes are independent given the random effects. Moreover, the longitudinal responses of each subject are assumed independent given the random effects. The posterior distribution is written as

$$p(\boldsymbol{\theta} \mid \mathbf{y}_i, T_i, \delta_i) \propto \prod_{j=1}^{n_i} p(y_{ij} \mid \mathbf{b}_i, \boldsymbol{\theta}_y) p\{T_i, \delta_i \mid \eta_i(T_i), \boldsymbol{\theta}_s\} p(\mathbf{b}_i \mid \boldsymbol{\theta}_y) p(\boldsymbol{\theta}_y) p(\boldsymbol{\theta}_s),$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_s^\top, \boldsymbol{\theta}_y^\top)^\top$ is the parameter vector for the survival and the longitudinal outcomes respectively. The likelihood contribution of the survival model is given in Web Appendix C in the Web-based Supplementary Materials. More details can be found in Rizopoulos (2012).

3.1 Bayesian P-splines

The Bayesian P-splines approach was first introduced by Lang and Brezger (2004). In our case, the smoothness of functions $\lambda(t)$ and $h_0(t)$ is controlled by the following priors for the coefficient that links the longitudinal and the survival outcomes $\boldsymbol{\alpha}$ and the coefficient of the baseline hazard $\boldsymbol{\gamma}_{h_0}$:

$$\begin{aligned}\boldsymbol{\alpha} \mid \tau_\alpha &\sim N_L(\mathbf{0}, \tau_\alpha \mathbf{M}_\alpha) \quad \text{and} \quad \tau_\alpha \sim \text{Gamma}(c_1, c_2), \\ \boldsymbol{\gamma}_{h_0} \mid \tau_{\gamma_{h_0}} &\sim N_U(\mathbf{0}, \tau_{\gamma_{h_0}} \mathbf{M}_{\gamma_{h_0}}) \quad \text{and} \quad \tau_{\gamma_{h_0}} \sim \text{Gamma}(f_1, f_2),\end{aligned}$$

where \mathbf{M}_α , $\mathbf{M}_{\gamma_{h_0}}$ are the penalty matrices. In particular, $\mathbf{M}_\alpha = \mathbf{M}_{\gamma_{h_0}} = \mathbf{D}_r^\top \mathbf{D}_r + 10^{-6} \mathbf{I}$, where \mathbf{D}_r is a r -th order difference matrix. The scaled identity matrix \mathbf{I} ensures a positive definite variance-covariance matrix. As described in the literature (Reinsch, 1967; Eilers and Marx, 1996), a common choice for the penalty matrices is to assume a second order penalty. The amount of smoothness is controlled by the variance parameters $\tau_{\gamma_{h_0}}$ and τ_α , where hyperpriors are assigned. A usual recommendation is to set c_1 and f_1 equal to 1 and c_2 and f_2 equal to a small number. Alternative specifications of these hyperpriors can be found in Jullion and Lambert (2007).

4. Measuring Predictive Performance

As motivated in Section 1, it is important for physicians to have a prognostic tool for planning next interventions. To assess the predictive performance of the VCJM and to compare it to the CCJM, we focus on discrimination and calibration. Specifically, discrimination is how well can the model discriminate between patients who will experience the event from patients

who will not (Pencina et al., 2008), whereas calibration is how well the model predicts the observed event rates (Schemper and Henderson, 2000).

4.1 Discrimination

A key feature of our model is to distinguish between patients who are going to die or require a reoperation within a specific time frame from patients who will not. In particular, for a future patient l with SAG measurements $\tilde{\mathbf{y}}_l$ up to time point t , we are interested in investigating whether he will die or require a reoperation in the medically-relevant time frame $(t, t + \Delta t]$ within which the physician could intervene to improve survival. The survival/intervention-free probability of patient l within this interval is,

$$\pi_l(t, \Delta t) = \Pr(T_l^* \geq t + \Delta t \mid T_l^* > t, \tilde{\mathbf{y}}_l(t), \mathbf{D}_n),$$

where $\mathbf{D}_n = \{T_i, \delta_i, \mathbf{y}_i, i = 1, \dots, n\}$ denotes the sample on which the joint model was fitted. For the calculation of sensitivity and specificity, we have $\pi_l(t, \Delta t) \leq c$ if subject l died or required a reoperation and $\pi_l(t, \Delta t) \geq c$ if he did not experience death/reoperation, for a specific $c \in [0, 1]$. In particular, we can define sensitivity and specificity as

$$\Pr\{\pi_l(t, \Delta t) \leq c \mid T_l^* \in (t, t + \Delta t]\} \quad \text{and} \quad \Pr\{\pi_l(t, \Delta t) > c \mid T_l^* > t + \Delta t\},$$

respectively. Using the area under the receiver operating characteristic curve (AUC) we can assess the discriminative capability of the model. In particular, given a randomly chosen pair of patients (l_1, l_2) , we have

$$\text{AUC}(t, \Delta t) = \Pr[\pi_{l_1}(t, \Delta t) < \pi_{l_2}(t, \Delta t) \mid \{T_{l_1}^* \in (t, t + \Delta t]\} \cap \{T_{l_2}^* > t + \Delta t\}].$$

If patient l_1 experiences death/reoperation within the relevant time frame whereas patient l_2 does not, then we would expect the VCJM to assign higher survival/intervention-free probability during the period $(t, t + \Delta t]$ for the patient that did not experience death/reoperation. In the cardio data set, the values of time-to-death/reoperation are not fully observed for

all patients. To account for this, the estimation of $\text{AUC}(t, \Delta t)$ is based on the following decomposition

$$\widehat{\text{AUC}}(t, \Delta t) = \sum_{w=1}^4 \widehat{\text{AUC}}_w(t, \Delta t),$$

which include the following pairs of patients

$$\begin{aligned} \Omega_{l_1 l_2}^{(1)}(t) &= [\{T_{l_1} \in (t, t + \Delta t)\} \cap \{\delta_{l_1} = 1\}] \cap \{T_{l_2} > t + \Delta t\}, \\ \Omega_{l_1 l_2}^{(2)}(t) &= [\{T_{l_1} \in (t, t + \Delta t)\} \cap \{\delta_{l_1} = 0\}] \cap \{T_{l_2} > t + \Delta t\}, \\ \Omega_{l_1 l_2}^{(3)}(t) &= [\{T_{l_1} \in (t, t + \Delta t)\} \cap \{\delta_{l_1} = 1\}] \cap [\{T_{l_1} < T_{l_2} \leq t + \Delta t\} \cap \{\delta_{l_2} = 0\}], \\ \Omega_{l_1 l_2}^{(4)}(t) &= [\{T_{l_1} \in (t, t + \Delta t)\} \cap \{\delta_{l_1} = 0\}] \cap [\{T_{l_1} < T_{l_2} \leq t + \Delta t\} \cap \{\delta_{l_2} = 0\}]. \end{aligned}$$

$\widehat{\text{AUC}}_1(t, \Delta t)$ includes the pairs of patients who are comparable $\Omega_{l_1 l_2}^{(1)}(t)$ and $\sum_{w=2}^4 \widehat{\text{AUC}}_w(t, \Delta t)$ the pairs of patients who due to censoring cannot be compared $\{\Omega_{l_1 l_2}^{(w)}(t), w = 2, 3, 4\}$. Then, with $I(\cdot)$ the indicator function

$$\widehat{\text{AUC}}_w(t, \Delta t) = \frac{\sum_{l_1=1}^n \sum_{l_2=1; l_2 \neq l_1}^n I\{\hat{\pi}_{l_1}(t, \Delta t) < \hat{\pi}_{l_2}(t, \Delta t)\} \times I\{\Omega_{l_1 l_2}^{(w)}(t)\} \times \hat{K}_w}{\sum_{l_1=1}^n \sum_{l_2=1; l_2 \neq l_1}^n I\{\Omega_{l_1 l_2}^{(w)}(t)\} \times \hat{K}_w},$$

which is the proportion of concordant subjects out of the set of comparable subjects at time t . Specifically, if in a randomly selected pair of patients, the one with the higher event probability experiences the event and the one with the lower probability does not experience the event, then this pair is said to be a concordant pair. For $w = 1$, we have $\hat{K}_2 = 1$ because the pairs of patients are comparable. For $w = 2, 3, 4$, the $\widehat{\text{AUC}}_w(t, \Delta t)$ are weighted with the probability that the concordant subjects are comparable. In particular, $\hat{K}_2 = 1 - \hat{\pi}_{l_1}(t, \Delta t)$, $\hat{K}_3 = \hat{\pi}_{l_2}(t, \Delta t)$ and $\hat{K}_4 = \{1 - \hat{\pi}_{l_1}(t, \Delta t)\} \times \hat{\pi}_{l_2}(t, \Delta t)$

4.2 Calibration

To assess the accuracy of the model, we use the prediction error (PE). Using all available information for a particular patient l , we are interested in comparing the predicted probability of survival/intervention-free of this patient to the observed truth:

$$\text{PE}(t, \Delta t) = E[\{N_l(t + \Delta t) - \pi_l(t, \Delta t)\}^2]$$

where $N_l(t) = I(T_l^* > t)$ is the event status at time t . To account for censoring, the following estimate has been proposed by Henderson et al. (2002):

$$\widehat{\text{PE}}(t, \Delta t) = \{\mathcal{R}(t)\}^{-1} \sum_{l: T_l \geq t} \left\{ I(T_l > t + \Delta t) \{1 - \hat{\pi}_l(t, \Delta t)\}^2 + \delta_l I(T_l < t + \Delta t) \{0 - \hat{\pi}_l(t, \Delta t)\}^2 + (1 - \delta_l) I(T_l < t + \Delta t) [\hat{\pi}_l(T_l, \Delta t) \{1 - \hat{\pi}_l(t, \Delta t)\}^2 + \{1 - \hat{\pi}_l(T_l, \Delta t)\} \{0 - \hat{\pi}_l(t, \Delta t)\}^2] \right\}$$

where $\mathcal{R}(t)$ denotes the number of subjects at risk at t . The term $I(T_l > t + \Delta t) \{1 - \hat{\pi}_l(t, \Delta t)\}^2$ refers to patients who were alive after $t + \Delta t$ and $\delta_l I(T_l < t + \Delta t) \{0 - \hat{\pi}_l(t, \Delta t)\}^2$ to patients who died before $t + \Delta t$. The remaining term refers to patients who were censored in the interval $[t, t + \Delta t]$.

5. Analysis of the Cardio Data Set

In this section we present the analysis of the cardio data introduced in Section 1. Our primary focus is to investigate the association between SAG with time-to-death/reoperation. In Web Figure 1, the evolution of the SAG for 12 randomly selected patients is presented, where it is shown that most of these patients have non-linear profiles. Therefore, we assumed a linear mixed-effects submodel including natural cubic splines for time. The DIC criterion indicated that the model assuming one internal knot at 5.02 years (corresponding to 50% of the observed follow-up times) in both the fixed- and random-effects parts had a better fit. Furthermore, we corrected for gender. The mixed-effects submodel for the SAG is

$$y_i(t) = \eta_i(t) + \epsilon_i(t) = \beta_0 + \beta_1 g_i + \sum_{v=1}^2 \beta_{v+1} ns(t; v) + b_{0i} + \sum_{v=1}^2 b_{vi} ns(t; v) + \epsilon_i(t),$$

where $y_i(t)$ are the measurements of SAG, $ns(\cdot)$ denotes the natural cubic splines, $\epsilon_i(t) \sim N(0, \sigma)$ and $\mathbf{b}_i \sim N_b(\mathbf{0}, \mathbf{\Sigma}_b^2)$.

To investigate the association between SAG and survival, we postulated the VCJM. Motivated by previous work (Andrinopoulou et al., 2015b), where different features of the SAG were found to have an influence on survival, we assumed the value and the slope of the

longitudinal outcome to be associated with death/reoperation. Furthermore, we corrected for gender. Specifically, the survival submodel takes the form

$$\begin{aligned} h_i(t, \boldsymbol{\theta}_s) &= \exp \left\{ \sum_{u=1}^U \gamma_{h_0,u} B_u(t) \right\} \exp \left\{ \gamma g_i + \lambda_1(t) \eta_i(t) + \lambda_2(t) \frac{d\eta_i(t)}{dt} \right\} \\ &= \exp \left\{ \sum_{u=1}^U \gamma_{h_0,u} B_u(t) \right\} \exp \left\{ \gamma g_i + \sum_{\ell=1}^L \alpha_{1\ell} B_\ell(t) \eta_i(t) + \sum_{\ell=1}^L \alpha_{2\ell} B_\ell(t) \frac{d\eta_i(t)}{dt} \right\}, \end{aligned}$$

where $\alpha_{1\ell}$ and $\alpha_{2\ell}$ are the coefficients that link the longitudinal and survival processes, γ is the coefficient for gender and $\gamma_{h_0,\ell}$ are the baseline hazard coefficients. We assumed in both cases quadratic B-splines basis $B_\ell(t) = B_u(t)$ with 8 equally distance internal knots ranging from zero until 20.1 years.

As described in Section 3, for the P-splines approach we assumed normal priors for the time-varying coefficients $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2\}$ and the baseline hazard coefficients $\boldsymbol{\gamma}_{h_0}$ and gamma hyperpriors for τ_α and $\tau_{\gamma_{h_0}}$ where we took $c_1 = f_1 = 1$ and $c_2 = f_2 = 0.005$. For the rest of the parameters standard noninformative priors were used. For the coefficients of the longitudinal outcome $\boldsymbol{\beta}$ and the survival coefficient γ normal priors were taken with mean zero and large variance. For the variance-covariance matrix of the random effects $\boldsymbol{\Sigma}_b$ we assumed inverse Wishart prior with an identity scale matrix and degrees of freedom equal to the total number of the random effects. For the precision parameter of the longitudinal outcome a gamma prior was taken with parameters that were based on the separate analysis of the outcome. We ran the MCMC with a single chain for 150,000 iterations, with a burn-in of 50,000 and thinning of 2.

In Figure 2 we present the mean estimates and the credible intervals of the estimated $\lambda(t)$ for the value and the slope association parameters, respectively. The grey solid lines represent the coefficient when CCJM was assumed. The CCJM takes the form,

$$h_i(t, \boldsymbol{\theta}_s) = \exp \left\{ \sum_{u=1}^U \gamma_{h_0,u} B_u(t) \right\} \exp \left\{ \gamma g_i + \alpha_1 \eta_i(t) + \alpha_2 \frac{d\eta_i(t)}{dt} \right\},$$

where we use P-splines for the baseline hazard for a fair comparison. We observe that the effect of the SAG on survival seems to slightly increase over time, however this effect is not

strong. The effect of the slope of SAG on survival appears to increase linearly with time. Specifically, at the beginning of the study ($t = 5$), for patients having the same gender and level of SAG, the log hazard ratio for one mmHg increase in the current slope of the SAG is 3.26. However, at the end of the study ($t = 15$) this effect increases to 7.03. Further results of the models including the posterior estimates, DIC and trace plots are presented in Web Tables 1 and 2 and Web Figures 2 and 3 followed by a discussion (Web Appendix B in the Web-based supplementary materials). In Web Figure 4, we present prediction plots for a 46 year old male. Every time the patient visits the hospital his survival/intervention-free probabilities are updated. Specifically, five years after his first visit his survival/intervention-free probability is 0.7, while five years after his last visit this probability is 0.3. The evolution of the SAG for the specific patient is presented in Web Figure 5.

[Figure 2 about here.]

To investigate whether the proposed VCJM improves dynamic predictions, we compared the VCJM with the CCJM based on the $AUC(t, \Delta t)$ and $PE(t, \Delta t)$ measures introduced in Section 4. Corrected estimates of these measures were obtained using an internal validation procedure. We performed a 5-fold cross-validation by splitting our data set in five subsets, fitting each time the model in four of the subsets and calculating the accuracy measures in the subset that was excluded. This cross-validation procedure was replicated 100 times. The calculation of $AUC(t, \Delta t)$ and $PE(t, \Delta t)$ was performed for the follow-up times $\{t = 5.5, 7.5, 9.5\}$ whereas $\Delta t = 2$. The results are presented in Figure 3 where boxplots of 100 cross validations are shown. We obtained better discriminative capability (AUC) and predictive accuracy (PE) from the VCJM model compared to the CCJM in all cases. Specifically, we observe most of the time higher AUC values and lower PE values for the VCJM. These differences seem to be more profound for later follow-up times.

[Figure 3 about here.]

6. Simulation study

6.1 Design

We performed a series of simulations to evaluate the performance of the VCJM. We simulated 400 patients with maximum number of repeated measurements equal to 10. We assumed one longitudinal and one survival outcome as in the analysis of the cardio data set. For the continuous longitudinal outcome, we investigated the following linear mixed-effects model

$$y_i(t) = \eta_i(t) + \epsilon_i(t) = \beta_0 + \beta_1 g_i + \beta_2 t + b_{0i} + b_{1i} t + \epsilon_i(t),$$

where $\epsilon_i \sim N(0, \sigma_y^2)$ and $\mathbf{b} = (b_{0i}, b_{1i}) \sim N_2(\mathbf{0}, \Sigma_b^2)$. For simplicity, we adopted a linear effect of time for both the fixed and the random part, and corrected for gender. Time t was simulated from a uniform distribution between zero and 19.5. For the survival part, we investigated the following scenarios.

For **Scenario I**, we postulated the following model:

$$h_i(t) = h_0(t) \exp \left\{ \gamma g_i + \lambda(t) \eta_i(t) \right\} = h_0(t) \exp \left\{ \gamma g_i + \sum_{\ell=1}^L \alpha_\ell B_\ell(t) \eta_i(t) \right\},$$

where $B_\ell(t)$ denotes the ℓ -th basis function of a B-spline where the knots were placed at fixed time points. We assumed time-varying effect of the underlying value for the association parameter and corrected for gender. The baseline risk was simulated from a Weibull distribution $h_0(t) = \xi t^{\xi-1}$. For the simulation of the censoring times, an exponential censoring distribution was chosen with mean μ_c , so that the censoring rate was between 40% and 60%. Under this setting we simulated 200 data sets assuming two scenarios: **Scenario Ia** which assumed a linear evolution for $\lambda(t)$, and **Scenario Ib** which assumed a non-linear evolution.

For **Scenario II**, the survival submodel takes the form,

$$h_i(t) = h_0(t) \exp \left\{ \gamma g_i + \alpha \eta_i(t) \right\},$$

where a constant effect for the α coefficient was assumed.

Finally, for **Scenario IIIa** and **Scenario IIIb**, the survival submodel takes the form,

$$h_i(t) = h_0(t) \exp \left\{ \gamma g_i + \alpha_1 + \alpha_2 \eta_i(t) + \alpha_3 \eta_i^2(t) \right\},$$

and

$$h_i(t) = h_0(t) \exp \left\{ \gamma g_i + \alpha_1 + \alpha_2 \eta_i(t) + \alpha_3 \eta_i^2(t) + \alpha_4 \eta_i^3(t) + \alpha_5 \eta_i^4(t) \right\},$$

respectively. In this case, we assumed polynomial effects for the association parameters.

More details about the parameters of the models are presented in Web Table 3.

6.2 Analysis and Results

To evaluate the performance of the proposed model, we fit the VCJM and the CCJM with the same specification as the one presented in Section 6.1. To mimic the analysis of the cardio data set, the joint models were fitted using P-splines for the baseline hazard, rather than the Weibull one. Web Figures 6 and 7, illustrate the true and estimated $\lambda(t)$ function when simulating from the linear Scenario Ia (left panel) and the non-linear Scenario Ib (right panel) and when fitting the VCJM. In particular, in Web Figure 6 we assumed 8 internal knots for the simulation and the fitting part while in Web Figure 7 we assumed 20. Web Figure 8 illustrates the true and estimated $\lambda(t)$ function for Scenario II when fitting the VCJM assuming 8 internal knots. Web Figures 9 and 10 present the results when simulating for Scenario IIIa and Scenario IIIb respectively and when fitting the VCJM with 8 internal knots. Overall our model successfully recovers the true $\lambda(t)$. The reason for observing greater variability at $t < 5$ and $t > 15$ is that relatively few events are observed in these regions.

To further evaluate whether the VCJM produces predictions of better quality than the CCJM, we performed an external validation procedure. For each simulated data set from scenarios Ia, Ib and II, we randomly excluded 200 patients. Using the remaining patients, we fitted the VCJM and the CCJM and computed the $AUC(t, \Delta t)$ and $PE(t, \Delta t)$ for the 200 patients that were initially excluded. These measures were calculated at follow-up times $\{t = 5.5, 7.5, 9.5\}$ using $\Delta t = 2$. Figures 4 and 5 present boxplots with the results under Scenarios Ia and Ib, respectively. It can be seen that, overall, the VCJM performs better

than the CCJM. Most of the time we observe a higher AUC value and a lower PE value for the VCJM. Smaller differences are obtained for Scenario Ib for the AUC at $t = 5.5$ and the PE at $t = 9.5$. This is explained by the fact that not enough events were observed at the specific time periods. Figure 6, which presents boxplots for Scenario II, suggests that the VCJM provides accurate predictions even if the data is generated with a constant effect for the coefficient that links the longitudinal and survival outcomes. In all cases, we obtain AUC and PE values that are similar in both the VCJM and the CCJM.

[Figure 4 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

7. Discussion

Motivated by the fact that the human tissue valves degenerate and hence the effect of SAG on survival could change over time, we developed a VCJM using P-splines. We used P-splines for the logarithm of the baseline hazard. We showed that even when the data is generated with a constant effect for the coefficient that links the longitudinal and survival outcome, the VCJM performs equal or better than the CCJM. In this paper we followed the Bayesian framework and therefore, extensions to more complex situations are comparably easy. A further advantage of the Bayesian approach, is that it allows us to automatically estimate the smoothing parameter in the P-splines approach by assigning a prior to it.

Further extensions to improve the VCJM can be considered for future work. In this paper we used only one longitudinal and one survival outcome. Extension to multiple longitudinal outcomes might be useful since other longitudinal outcomes that are related to the heart value may have an influence on survival. Finally, death and reoperation are competing risks and therefore a competing risk survival submodel would be more appropriate.

SUPPLEMENTARY MATERIALS

Web Figures 1-10, Web Tables 1-3 and discussion referenced in Sections 3, 5 and 6, are available with this paper at the Biometrics website on Wiley Online Library.

REFERENCES

- Andrinopoulou, E.-R. and Rizopoulos, D. (2016). Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. DOI: 10.1002/sim.7027.
- Andrinopoulou, E.-R., Rizopoulos, D., Geleijnse, M. L., Lesaffre, E., Bogers, A. J., and Takkenberg, J. J. (2015a). Dynamic prediction of outcome for patients with severe aortic stenosis: application of joint models for longitudinal and time-to-event data. *BMC cardiovascular disorders* **15**, 1.
- Andrinopoulou, E.-R., Rizopoulos, D., Jin, R., Bogers, A. J., Lesaffre, E., and Takkenberg, J. J. (2012). An introduction to mixed models and joint modeling: analysis of valve function over time. *The Annals of thoracic surgery* **93**, 1765–1772.
- Andrinopoulou, E.-R., Rizopoulos, D., Takkenberg, J. J., and Lesaffre, E. (2014). Joint modeling of two longitudinal outcomes and competing risk data. *Statistics in medicine* **33**, 3167–3178.
- Andrinopoulou, E.-R., Rizopoulos, D., Takkenberg, J. J., and Lesaffre, E. (2015b). Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. DOI: 10.1177/0962280215588340.
- Bekkers, J. A., Klieverik, L. M., Raap, G. B., Takkenberg, J. J., and Bogers, A. J. (2011). Re-operations for aortic allograft root failure: experience from a 21-year single-center prospective follow-up study. *European Journal of Cardio-Thoracic Surgery* **40**, 35–42.
- Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing using b-splines and penalized likelihood (with comments and rejoinder). *Statistical Science* **11**, 1200–1224.

- Eilers, P. H. C., Marx, B. D., and Durbán, M. (2015). Twenty years of p-splines. *SORT-Statistics and Operations Research Transactions* **39**, 149–186.
- Garre, F. G., Zwinderman, A. H., Geskus, R. B., and Sijpkens, Y. W. (2008). A joint latent class changepoint model to improve the prediction of time to graft failure. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **171**, 299–308.
- Henderson, R., Diggle, P., and Dobson, A. (2002). Identification and efficacy of longitudinal markers for survival. *Biostatistics* **3**, 33–50.
- Jullion, A. and Lambert, P. (2007). Robust specification of the roughness penalty prior distribution in spatially adaptive bayesian p-splines models. *Computational statistics & data analysis* **51**, 2542–2558.
- Lang, S. and Brezger, A. (2004). Bayesian p-splines. *Journal of computational and graphical statistics* **13**, 183–212.
- Nan, B., Lin, X., Lisabeth, L. D., and Harlow, S. D. (2005). A varying-coefficient cox model for the effect of age at a marker event on age at menopause. *Biometrics* **61**, 576–583.
- Pencina, M. J., D’Agostino, R. B., and Vasan, R. S. (2008). Evaluating the added predictive ability of a new marker: from area under the roc curve to reclassification and beyond. *Statistics in medicine* **27**, 157–172.
- Perperoglou, A. (2014). Cox models with dynamic ridge penalties on time-varying effects of the covariates. *Statistics in medicine* **33**, 170–180.
- Proust-Lima, C. and Taylor, J. M. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modeling approach. *Biostatistics* **10**, 535–549.
- Reinsch, C. H. (1967). Smoothing by spline functions. *Numerische mathematik* **10**, 177–183.
- Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* **67**, 819–829.

- Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. Chapman and Hall/CRC Biostatistics Series, Boca Raton.
- Rizopoulos, D. and Ghosh, P. (2011). A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in medicine* **30**, 1366–1380.
- Rizopoulos, D., Hatfield, L. A., Carlin, B. P., and Takkenberg, J. J. (2014). Combining dynamic predictions from joint models for longitudinal and time-to-event data using bayesian model averaging. *Journal of the American Statistical Association* **109**, 1385–1397.
- Rizopoulos, D., Taylor, J. M., Van Rosmalen, J., Steyerberg, E. W., and Takkenberg, J. J. (2015). Personalized screening intervals for biomarkers using joint models for longitudinal and survival data. 10.1093/biostatistics/kxv031.
- Schemper, M. and Henderson, R. (2000). Predictive accuracy and explained variation in cox regression. *Biometrics* **56**, 249–255.
- Song, X. and Wang, C. (2008). Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients. *Biometrics* **64**, 557–566.
- Taylor, J. M., Yu, M., and Sandler, H. M. (2005). Individualized predictions of disease progression following radiation therapy for prostate cancer. *Journal of clinical oncology* **23**, 816–825.
- Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* **14**, 809–834.
- Yu, M., Taylor, J. M. G., and Sandler, H. M. (2008). Individual prediction in prostate cancer studies using a joint longitudinal survival–cure model. *Journal of the American Statistical Association* **103**, 178–187.

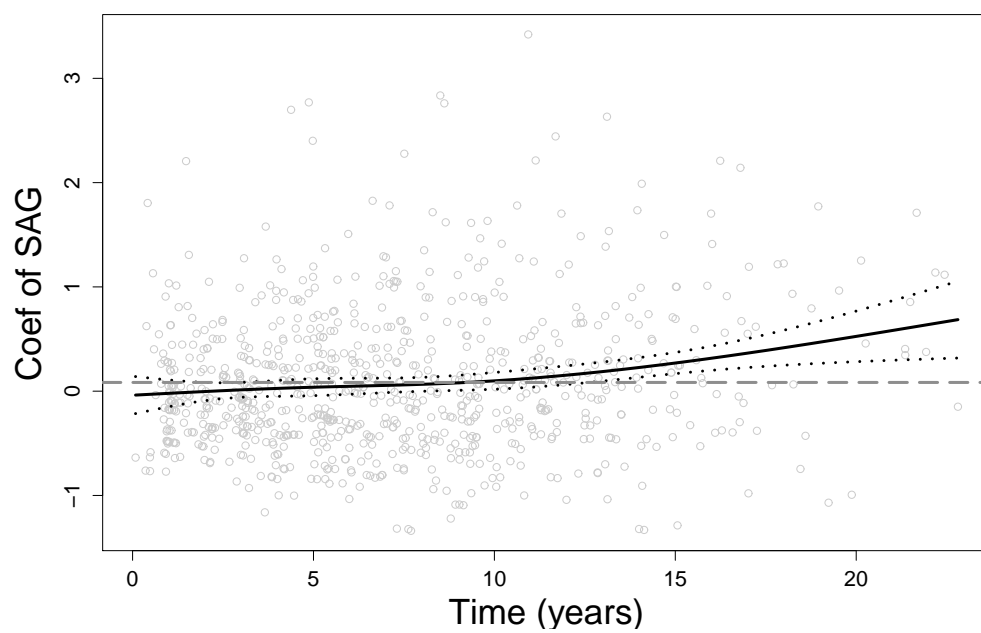


Figure 1. Time-varying coefficient of the SAG, using Schoenfeld residuals. The solid line represents the mean estimate while the dotted lines the corresponding confidence interval. The dashed grey line represents the coefficient of SAG.

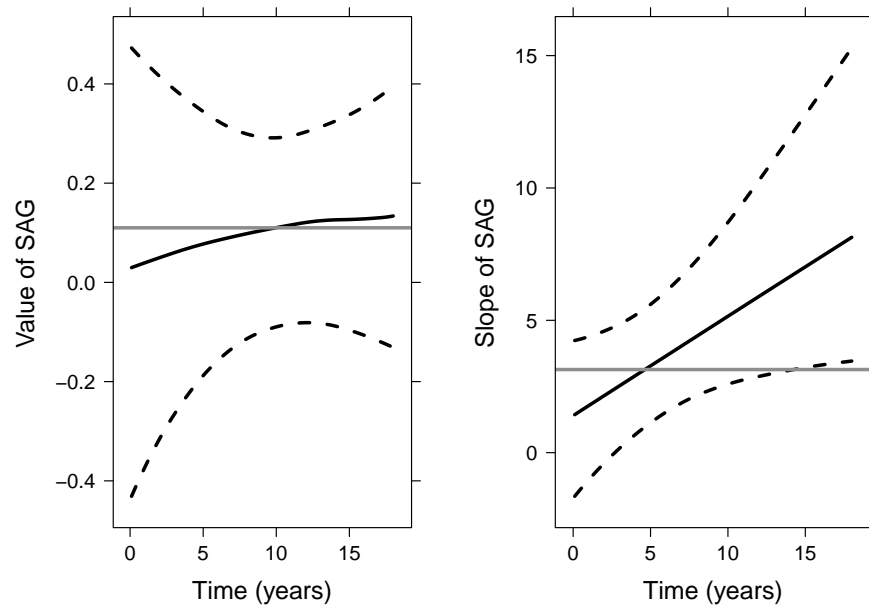


Figure 2. With black lines the mean estimates (solid line) and the credible intervals (dashed lines) of $\lambda_1(t)$ and $\lambda_2(t)$ functions corresponding to the value and slope association parameters are presented. The horizontal grey solid lines represent the mean estimates of the CCJM.

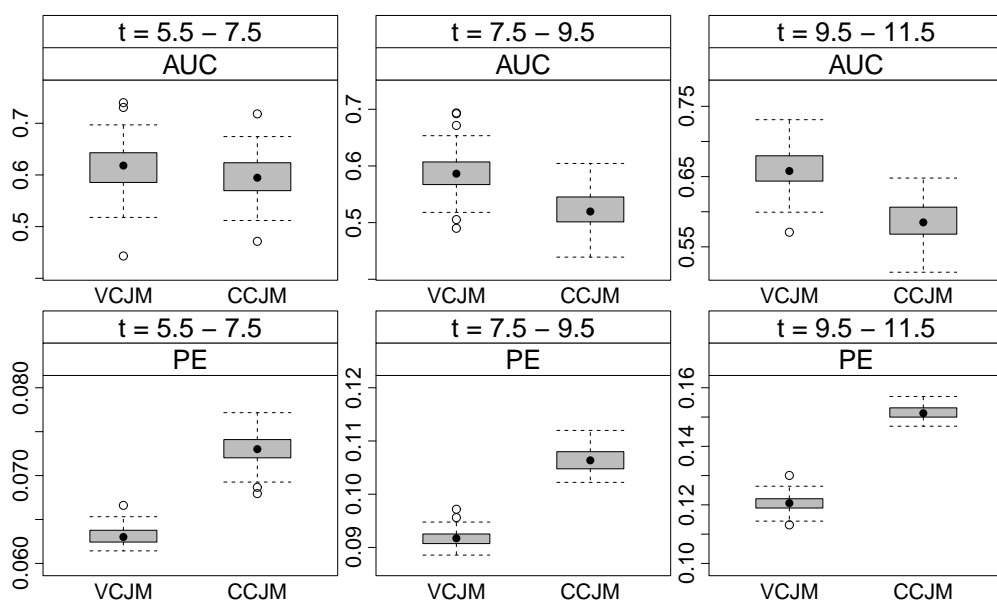


Figure 3. Boxplots of AUC and PE when assuming the VCJM and CCJM at different time points with $\Delta t = 2$. Internal validation.

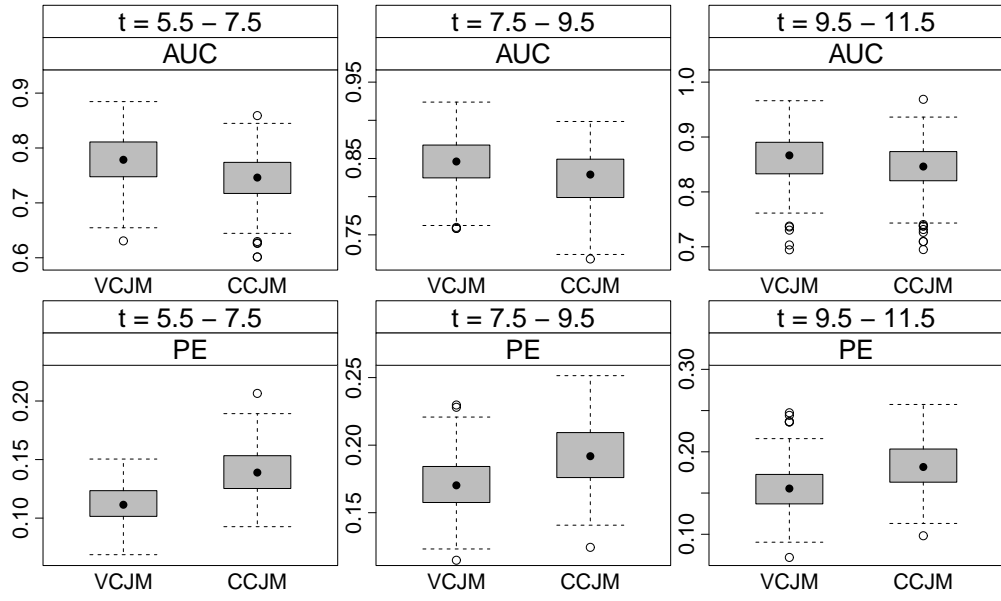


Figure 4. Boxplots of AUC and PE when assuming the VCJM and the CCJM at different time points with $\Delta t = 2$. For the simulation of the data scenario Ia was used. External validation.

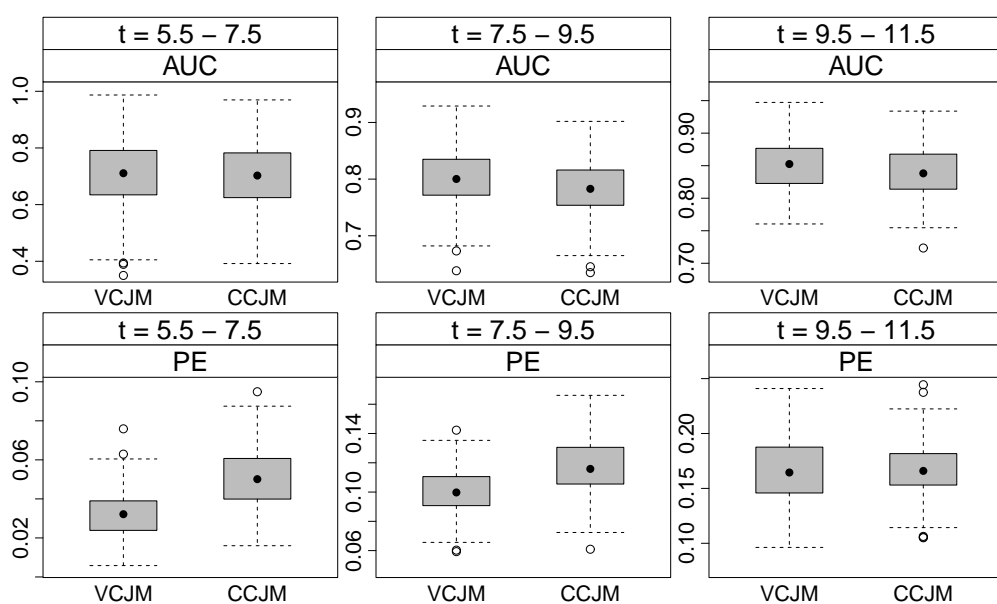


Figure 5. Boxplots of AUC and PE when assuming the VCJM and the CCJM at different time points with $\Delta t = 2$. For the simulation of the data scenario Ib was used. External validation.

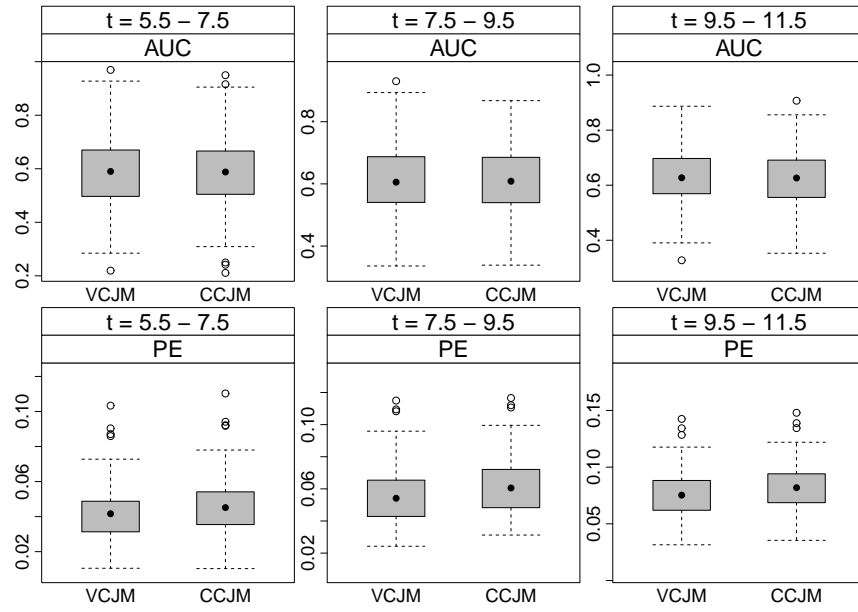


Figure 6. Boxplots of AUC and PE when assuming the VCJM and the CCJM at different time points with $\Delta t = 2$. For the simulation of the data scenario II was used. External validation.