

# Dynamic prediction of mortality amongst patients in intensive care using the sequential organ failure assessment (SOFA) score: A joint competing risk survival and longitudinal modelling approach.

J.Z.Musoro<sup>\*,1</sup>, A.H. Zwinderman<sup>1</sup>, R. Bosman<sup>2</sup>, A. Abu-Hanna<sup>3</sup> and R.B. Geskus<sup>1,4,5</sup>

<sup>1</sup> Department of Clinical Epidemiology, Biostatistics and Bioinformatics Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands

<sup>2</sup> Department of Intensive Care, Onze Lieve Vrouwe Gasthuis, Oosterpark 9, 1091 AC Amsterdam, The Netherlands

<sup>3</sup> Department of Medical Informatics, Academic Medical Center, Universiteit van Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

<sup>4</sup> Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom, Oxford University Clinical Research Unit, Wellcome Trust

<sup>5</sup> Major Overseas Programme, Ho Chi Minh City, Viet Nam and Department of Clinical Epidemiology, Biostatistics and Bioinformatics

## Abstract

In intensive care units (ICU), besides routinely collected admission data, a daily monitoring of organ dysfunction using scoring systems such as the Sequential Organ Failure Assessment (SOFA) score has become practice. Such updated information is valuable in making accurate predictions of patients survival. Few prediction models that incorporate this updated information have been reported. We used follow-up data of ICU patients who either died or were discharged at the end of hospital stay, without censored cases. We propose a joint model comprising a linear mixed effects submodel for the development of longitudinal SOFA scores, and a proportional subdistribution hazards submodel for death as end point with discharge as competing risk. The two parts are linked by shared latent terms. Since there was no censoring, it was straightforward to fit our joint model using available software. We compared predictive values, based on the Brier score and the Area Under the Receiver Operating Characteristic Curve (AUC), from our model with those obtained from an earlier modelling approach by Toma et al. (2007) which relied on patterns discovered in the SOFA scores over a given period of time.

Keywords: Competing risks; Dynamic prediction; Joint models; Longitudinal data; Survival data.

## 1 Introduction

In intensive care units (ICU), the prediction of survival status amongst patients at the end of hospital stay commonly rely solely on data collected during the first 24 hours of admission to the ICU. Besides this admission related data, a daily monitoring of information on patients'

organ dysfunction using different scoring systems such as the Sequential Organ Failure Assessment (SOFA) score (Vincent and Ferreira, 2000) has recently gained attention. It is important to be able to update prognosis using the most recent (daily) information. Hence models that make use of these updated scores to characterize the evolving mortality prediction during the ICU stay are advocated.

The SOFA scores comprise 6 individual organ system scores, each ranging from 0 to 4 (Vincent and Ferreira, 2000). In the literature, in order to predict mortality amongst ICU patients, most authors either used; (i) only the SOFA scores at admission or at a fixed time point after admission (Khwannimit, 2007; Ho et al., 2007), (ii) aggregates such as the mean or the maximum SOFA score in a prespecified time interval (Ho et al., 2007; Kajdacsy-Balla et al., 2005) or (iii) models which rely on patterns discovered in the SOFA scores over a given period of time (Toma et al., 2007, 2008). Some studies combined SOFA score aggregates with other baseline covariates (Kajdacsy-Balla et al., 2005; Cabr et al., 2005), and sometimes only one of the 6 components of the SOFA score were used (Khwannimit, 2007; Moreno et al., 1999). For a review on the use of SOFA-based models for predicting mortality amongst ICU patients see Minne et al. (2008).

We looked at ICU data from the study by Toma et al. (2007). Patients' demographic information as well as physiological and laboratory information were collected at admission, and summarized as the severity-of-illness-score SAPS-II (hereafter SAPS in short). The SOFA score was used for daily monitoring of organ dysfunction. Using data of patients under follow-up at a given day, Toma et al. (2007) compared two logistic regression models with respect to predicting the probability of dying at the end of hospital stay; (i) a static model that included only the SAPS scores and (ii) a temporal model that included the SAPS scores and updated information from the SOFA scores called aligned episodes. Aligned episodes refer to patterns of SOFA score categories that are observed from the day at which prediction is made (for instance, day 3), down to the previous days (day 2 and day 1). First, three categories of the SOFA scores were created (low L, medium M, high H). If a patient recorded M, H and H on the first 3 days respectively, then his or her aligned episodes would be MH for day 2 and MHH for day 3. Thus, the possible number of aligned episodes for a given day  $k$  equals  $\sum_{j=1}^k 3^j$  which is large for day 5 (363) and huge for day 10 (88,572). For each day, the data was mined to identify the most common and most predictive aligned episodes. The selected episodes were then included as a categorical covariate in the model for the given day. Toma et al. (2007) derived the best aligned episodes only until day 5. Although a better predictive performance was observed with these temporal models (here-

after Toma approach), the fact that the SOFA scores were summarized in categories may have led to loss of information. Also, the model construction procedure became too computer intensive if we desired to search for aligned episodes beyond day 5. A remedy would be to rely on the observed best episodes until day 5 to make predictions for patients still at risk after day 5, even though this may be unreliable. We could also consider the past 5 days as time goes on. However, this is out of the scope of this paper.

As an alternative approach, we propose to use the joint modelling of survival and longitudinal data framework (Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997; Xu and Zeger, 2001; Tsiatis and Davidian, 2004; Rizopoulos, 2011; Njagi et al., 2013; Geskus, 2014; Musoro et al., 2015). The joint modelling approach uses the data more efficiently since the time-to-event and longitudinal subprocesses are modeled simultaneously. Furthermore, it may attain unbiased parameter estimates if dropout is informative. Finally, by modelling the longitudinal SOFA scores, we account for the fact that the SOFA scores may be measured with error. The fitted joint model can be used to calculate and dynamically update individual predictions using information that is available at new time points.

In recent years, the application of joint models has extended beyond the traditional single longitudinal response and single event type setting. For instance, our data presents a competing risks scenario where the occurrence of the event of interest, death in the hospital, is prevented by live discharge from hospital. We considered death in the hospital as endpoint since it allowed us to compare our findings with those from the Toma approach. Joint models for longitudinal data and competing risks event times have been addressed in the literature (Gueorguieva et al., 2012; Li et al., 2010; Elashoff et al., 2008). With respect to applications in an ICU setting, Deslandes and Chevret (2010) used a proportional subdistribution hazards submodel for competing risks in a joint modelling approach to estimate the effect of a treatment on SOFA scores. To the best of our knowledge, this modelling approach has not been previously used as a tool to perform dynamic predictions of the probability of dying amongst ICU patients.

We fitted a joint model for the time updated SOFA scores and mortality risk, while handling discharge as a competing event. Our joint model comprises the following two parts which are linked by shared latent terms. (i) A linear mixed effects submodel for the development of longitudinal SOFA scores and (ii) a proportional subdistribution hazards submodel for death in the hospital as end point with discharge as competing risk (Fine and Gray, 1999). The subdistribution hazard is the hazard that corresponds to the cause specific-cumulative incidence. As opposed to the commonly used proportional cause-specific hazards formula-

tion, it directly translates to event-specific survival probabilities. There was no censoring, nor late entry that induced left truncation in our data, making it easy to fit our joint model using the JM package(Rizopoulos, 2010) within the R statistical software(R Core Team, 2013). Since patients with the competing event remain in the risk set when estimating the subdistribution hazards, the prediction as implemented within JM(Rizopoulos, 2011) mimics the standard survival setting by calculating the probability given that you are alive, also including discharged individuals. But for the discharged individuals, it is implausible to calculate the probability of dying in the hospital after time of discharge, although we could also argue that their death probabilities are zero after time of discharge. Additional routines were developed to calculate the probability of dying in the hospital given that the patient is alive and has not been discharged.

We compared the performance of our joint model to that of the Toma approach for the first 20 days of ICU stay. Comparisons were based on Brier scores and the area under the receiver operating characteristic curve (AUC), which were calculated using the estimated probability of dying in the hospital from both the joint modelling and Toma approaches. Once we have obtained the individual probabilities, calculating the Brier scores and AUC was straightforward since there was no censored data.

The rest of this paper is organized as follows. First we present the study data, followed by a description of the submodels and their joint likelihood function. Next we present results from the joint model as well as from the dynamic predictions. We also compare the predictive performance of the joint model and the Toma approach.

## 2 Data

We looked at ICU data from the OLVG, a teaching hospital in Amsterdam. Patients were admitted between July 1998 and February 2007. The same data was used by Toma et al. (2007) except that we have additional data for patients seen between September 2005 and February 2007. In this paper, we considered 3697 patients who had been admitted for at least 24 hours in the ICU. SAPS scores were calculated for all patients within 24 hours of admission. Patients' organ dysfunction was monitored daily using the SOFA scores (integers ranging from 0 to 24). Larger SOFA scores correspond to worsening organ dysfunction. In the case where patients were readmitted to the ICU (which occurred in about 6% of the cases), only information from their last readmission was considered. At the end of follow-up, patients were either dead (797 deaths) or had been discharged alive from the hospital

(2900 discharged cases). Figure 1 shows the estimated cumulative incidence curves for the competing failure types. The number of SOFA score measurements within the ICU ranged from 1 to 85 (mean=10) and 1 to 144 (mean=8) measurements, respectively, for patients who died and who were discharged. The SOFA scores ranged from 2 to 21 (median=9) for those who died and 2 to 20 (median=7) for discharged cases. The SAPS score ranged from 22 to 111 (median = 57) for patients who died and from 9 to 96 (median=43) for those who were discharged. For an elaborate description of the data see Toma et al. (2007).

Figure 2 shows the longitudinal trajectories of the SOFA scores. For clarity, we show only the profiles for 500 randomly selected subjects. The lower panel of figure 2 illustrates the trajectory of a patient (Patient A), with a SAPS score of 46, who was discharged from the hospital after 17 days, and another (Patient B), with a SAPS score of 35, who died 16 days after admission. We observe that Patient A had a decreasing trajectory in his or her SOFA score levels and therefore is expected to have higher survival probabilities. Patient B on the other hand had an increasing pattern, which was indicative of a worsening condition.

## 3 Method

### 3.1 Model definition

For patient  $i$  ( $i = 1, \dots, n$ ), let  $Y_{ij}$  be the  $j^{th}$  SOFA score at time  $t_{ij}$ , ( $j = 1, 2, \dots, n_i$ ).  $Y_{ij}$  is only available while staying in the ICU. Let  $(T_i, E_i)$  represent the event data for patient  $i$ , where  $T_i$  is the event time and  $E_i$  is the event type. The number of competing events is equal to 2 for our data ( $l=1$  in case of death and  $l=2$  in case of discharge from hospital). If a competing event occurs, then we can say that the event of type  $l$  happens at time infinity, hence we can define:

$$T_i^l = \begin{cases} T_i & \text{if } E_i = l \\ \infty & \text{if } E_i \neq l \end{cases} \quad (1)$$

The joint model comprises the following two submodels.

(1) A linear mixed effects (LME) model specified as

$$Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + \alpha X_i + \epsilon_{ij} \quad (2)$$

where  $\beta_0$  is the intercept and  $\beta_1$  is the slope, and  $b_{0i}$  and  $b_{1i}$  are respectively the subject specific intercepts and slopes which are assumed to follow a zero-mean bivariate Gaussian

distribution with covariance matrix  $\Sigma$ .  $X_i$  is the baseline SAPS score with fixed effect  $\alpha$ . The error term  $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$  is assumed to be independent of  $b_{0i}$  and  $b_{1i}$ .

(2) A proportional subdistribution hazards model specified as

$$\begin{aligned}\lambda_i^l(t|X_i, Y_i^*(t)) &= \lim_{\Delta t \downarrow 0} \frac{\Pr(t \leq T_i^l < t + \Delta t | T_i^l \geq t, X_i, Y_i^*(t))}{\Delta t} \\ &= \lambda_0^l(t) \exp \left\{ \alpha^l X_i + \beta^l Y_i^*(t) \right\}\end{aligned}\quad (3)$$

where  $Y_i^*(t)$  is the expected SOFA score given  $(b_{0i}, b_{1i})$  for subject  $i$  at time  $t$  according to equation 2. Since there was no right censoring in our data, individuals who had the competing event remained in the risk set with a weight of 1 when estimating the subdistribution hazards.

### 3.2 The Likelihood Function

Let  $\mathbf{b}_i = (b_{0i}, b_{1i})$  and let  $\boldsymbol{\theta} = (\beta_0, \beta_1, \alpha, \sigma_\epsilon^2, \Sigma, \alpha^l, \beta^l)$  be a vector containing all parameters from (2) and (3). We postulate that the subprocesses represented by (2) and (3) are conditionally independent given  $\mathbf{b}_i$  and the covariate  $(X_i)$ . Consider death ( $l = 1$ ) to be the event of interest. The joint likelihood of the longitudinal and the competing risks submodel, based on all observed data, is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n \int_{\mathbf{b}_i} \left[ \left\{ \prod_{j=1}^{n_i} f_1(Y_{ij}|X_i, \mathbf{b}_i, \boldsymbol{\theta}) \right\} \times f_2(T_i^1|X_i, \mathbf{b}_i, \boldsymbol{\theta}) \times f_3(\mathbf{b}_i|\boldsymbol{\theta}) \right] d\mathbf{b}_i \quad (4)$$

$f_1(\cdot)$  is the univariate normal density of the SOFA score given  $\mathbf{b}_i$  and  $f_3(\cdot)$  is a bivariate normal density of the random effects.  $f_2(\cdot)$  is the subdistribution density, which can be written as

$$f_2(T_i^1|X_i, \mathbf{b}_i, \boldsymbol{\theta}) = \lambda_i^1(T_i|X_i, \mathbf{b}_i, \boldsymbol{\theta})^{I(l=1)} \times S_i^1(T_i|X_i, \mathbf{b}_i, \boldsymbol{\theta}) \quad (5)$$

$\lambda_i^1(\cdot)$  is given by (3) and  $S_i^1(\cdot)$  is the probability to be alive, also outside the hospital, which is defined as

$$S_i^1(T_i|X_i, \mathbf{b}_i, \boldsymbol{\theta}) = \exp \left\{ - \int_0^{T_i} \lambda_i^1(u|X_i, \mathbf{b}_i, \boldsymbol{\theta}) du \right\} \quad (6)$$

When evaluating (4), all patients who were discharged remain in the risk set forever.

### 3.3 Dynamic prediction of conditional survival probabilities

Using the fitted joint model, we wanted to calculate predictions for a new subject  $k$  who is alive in the hospital at  $t$  and has SOFA scores history  $\mathcal{D}(t) = \{Y_{kj}; 0 \leq t_{kj} \leq t\}$ . For any time  $u > t$  we were interested in the conditional probability  $\pi_k(u|t)$  that the new subject  $k$  will not die in the hospital at least up to  $u$ , given that he or she has not been discharged and is still alive at  $t$ . This can be written as

$$\pi_k(u|t) = \Pr(T_k^1 \geq u | T_k^1 \geq t, T_k^2 \geq t, \mathcal{D}(t), X_k) \quad (7)$$

Since the probability of dying in the hospital and the probability of discharge cannot both occur, then the probability of either occurring is the sum of the probabilities of each occurring. We can then rewrite (7) as

$$\begin{aligned} \Pr(T_k^1 \geq u | T_k^1 \geq t, T_k^2 \geq t, \mathcal{D}(t), X_k) = \\ \int \frac{1 - F_1(u|\mathbf{b}_k, X_k) - F_2(t|\mathbf{b}_k, X_k)}{1 - F_1(t|\mathbf{b}_k, X_k) - F_2(t|\mathbf{b}_k, X_k)} p(\mathbf{b}_k | T_k^1 \geq t, T_k^2 \geq t, \mathcal{D}(t), X_k) d\mathbf{b}_k \end{aligned} \quad (8)$$

where  $F_1(\cdot|\mathbf{b}_k, X_k)$  is the cumulative death probability, given the random effects of patient  $k$ .  $F_2(\cdot|\mathbf{b}_k, X_k)$  is the cumulative discharge probability. The density function  $p(\mathbf{b}_k | T_k^1 \geq t, T_k^2 \geq t, \mathcal{D}(t), X_k)$  is the distribution of the random effects given the SAPS score and observed SOFA scores up to day  $t$ , and given alive in the hospital:

$$\begin{aligned} p(\mathbf{b}_k | T_k^1 \geq t, T_k^2 \geq t, \mathcal{D}(t), X_k) = \\ \frac{L(\mathcal{D}(t)|\mathbf{b}_k, X_k) \times (1 - F_1(t|\mathbf{b}_k, X_k) - F_2(t|\mathbf{b}_k, X_k)) \times f_3(\mathbf{b}_k|\boldsymbol{\theta})}{\int L(\mathcal{D}(t)|\mathbf{b}_k, X_k) \times (1 - F_1(t|\mathbf{b}_k, X_k) - F_2(t|\mathbf{b}_k, X_k)) \times f_3(\mathbf{b}_k|\boldsymbol{\theta}) d\mathbf{b}_k} \end{aligned} \quad (9)$$

where  $L(\mathcal{D}(t)|\mathbf{b}_k, X_k)$  is the likelihood of  $\mathcal{D}(t)$  according to the linear-mixed effects submodel of the joint model given the random effects  $\mathbf{b}_k$ . Because the SOFA scores are independent given  $\mathbf{b}_k$ , this likelihood is a product of the likelihoods of each available SOFA score. This univariate likelihood is according to the normal distribution with mean  $(\beta_0 + b_{0k}) + (\beta_1 + b_{1k})t_{kj} + \alpha X_k$  and the residual standard deviation  $\sigma_\epsilon$  according to (2). Notice that  $1 - F_1(u|\mathbf{b}_k, X_k) = S_k^1(u|\mathbf{b}_k, X_k)$  from (6), suppressing  $\boldsymbol{\theta}$ . Now (8) can be written as

$$\begin{aligned} \Pr(T_k^1 \geq u | T_k^1 \geq t, T_k^2 \geq t, \mathcal{D}(t), X_k) = \\ \frac{\int_{\mathbf{b}_k} S_k^1(u|\mathbf{b}_k) \times L(\mathcal{D}(t)|\mathbf{b}_k) \times f_3(\mathbf{b}_k) d\mathbf{b}_k - \int_{\mathbf{b}_k} F_2(t|\mathbf{b}_k) \times L(\mathcal{D}(t)|\mathbf{b}_k) \times f_3(\mathbf{b}_k) d\mathbf{b}_k}{\int_{\mathbf{b}_k} S_k^1(t|\mathbf{b}_k) \times L(\mathcal{D}(t)|\mathbf{b}_k) \times f_3(\mathbf{b}_k) d\mathbf{b}_k - \int_{\mathbf{b}_k} F_2(t|\mathbf{b}_k) \times L(\mathcal{D}(t)|\mathbf{b}_k) \times f_3(\mathbf{b}_k) d\mathbf{b}_k} \end{aligned} \quad (10)$$



suppressing  $X_k$ . We used Gauss-Hermite quadrature to approximate the integrals in (10).

In order to compare predictive performance, a data set of 2683 patients who were seen between July 1998 and August 2005 was used to estimate the parameters in the joint model. The same data was used to search for the best aligned episodes for each of the first 5 days of ICU stay, as described by Toma et al. (2007). The performance was quantified using the test data of 1014 patients who were seen after August 2005 up till February 2007. The AUC was used to evaluate the discriminative performance, that is, how well patients who died were distinguished from those who were alive at the end of hospital stay, using the predicted probabilities. We also used the Brier score, i.e. the mean of the squared difference between the observed outcome and the predicted probability, which reflects both discrimination and calibration (the agreement between the observed outcome and the predictions). Both performance measures were evaluated for each of the first 20 days after entry into the ICU, using only patients remaining in the hospital up till the particular day. In the Toma approach, since the best aligned episodes were derived only until day 5, the predicted probabilities for patients who were followed-up till day 6 and onwards were based on their best aligned episodes at day 5.

## 4 Results

Parameter estimates from the joint model, with death as the event of interest, are shown in table 1. We provide estimates for the longitudinal and the proportional subdistribution hazards submodel. We observed an overall downward trend in the SOFA score trajectories ( $\beta_{time} = -0.3504$  with 95% CI: -0.3535,-0.3473). The probability of dying in the hospital was significantly higher for larger SOFA and SAPS scores. A unit increase in SOFA score was associated with an increase in the estimated hazard of death by  $\exp(0.3679) = 1.445$  (95% CI: 1.419, 1.471), after adjustment for the effect of the SAPS score. In figure 3, we illustrate how predicted survival probabilities for Patient A and B are dynamically updated based on new SOFA scores that became available every day. The plots show the observed and fitted SOFA score trajectories on the left, and the probability to be alive in the hospital on the right. Notice that we use different time scales for the SOFA scores and the survival probability. As expected, for Patient A who had lower SOFA scores, higher survival probabilities were observed compared to Patient B who showed worsening conditions (increasing SOFA scores) over time.



In figure 4 we present the predictive values (Brier scores and AUC's) from the joint modelling and the Toma approach. First, for days 1 to 5, we observe that both approaches performed similarly well. Then for day 6 and later, for both the Brier score and AUC, the joint modelling approach outperformed the Toma approach. The Toma approach did not use updated information after day 5, and apparently it was insufficient to rely on the best aligned episodes at day 5 to make accurate predictions after day 5.

## 5 Discussion

When analyzing IC data to predict the risk of dying in the hospital, besides the commonly used admission related information, it is important to also exploit information contained in daily SOFA scores. This helps to improve prognosis on the basis of the latest available information (Toma et al., 2007). Few models have been reported in the literature that incorporate the updated SOFA score trajectories to make new predictions despite the fact that the SOFA scores are routinely collected nowadays.

We looked at ICU data that was previously studied by Toma et al. (2007). We proposed to use the joint modelling framework to efficiently study the relationship between the SOFA score development and the probability of dying at the end of hospital stay. Discharge from the hospital was seen as a competing event. Our joint model comprised a linear mixed effects submodel for the development of longitudinal SOFA scores and a proportional subdistribution hazards submodel for death in the hospital. Both approaches performed similarly well during the first 5 days of hospital stay. Probably a more elaborated fixed effects structure in the longitudinal submodel (i.e including more covariates that are prognostic of patients' state of health), would have yielded a better discriminative ability of the SOFA scores. However, compared to the Toma approach, it was more straightforward to fit our joint model using freely available software.

Cause-specific hazard submodels have been commonly used in the literature because the time-varying variables can be handled in a standard way (Gueorguieva et al., 2012; Li et al., 2010; Elashoff et al., 2008). In our data, all patients were observed to have died or discharged. This made it easy to estimate a subdistribution hazard model with time-fixed covariates by having patients that are discharged stay in the risk set with a weight of 1; we did not have to use weighted risk sets (Geskus, 2011). Furthermore, there were no censored cases, making it straightforward to fit our model within the JM package (Rizopoulos, 2010). Patients' current fitted SOFA score values from the random effects model were included as

a time-varying covariate. Inclusion of time-varying covariates in a subdistribution hazard model is not straightforward, but Deslandes and Chevret (2010) showed in a simulation study that this approach yielded accurate estimates. It remains a subject for discussion whether or not it is correct to use a subdistribution hazard submodel with time-varying covariates (Beyersmann and Schumacher, 2008). One may question the interpretability of a fitted value of the longitudinal outcome for subjects who remained in the risk set in the analysis despite already experiencing the competing event. With our data for instance, we could argue that with a non-lethal competing event like discharge, the SOFA score values are measurable in theory although there is no guarantee that the true SOFA score values after discharge are the same as what is predicted. But judging by the predictions from our joint model, this seems to work well at least for our data. On the other hand, if we were interested in discharge, then we can not measure SOFA scores for patients who died and so their predicted values will be highly speculative. Beyersmann and Schumacher (2008) suggested to carry forward the last observed value of the time-varying covariate when the competing event has occurred. Since time-varying covariates are handled in the standard way in a cause-specific hazards model, for future research it would be interesting to compare predictions from a joint model that uses a subdistribution hazards submodel to one with a cause-specific hazards submodel.

For the dynamic predictions within JM, it was straightforward to calculate the probability that a patient will die in the hospital at certain moments in time given his or her history of SOFA scores, and given her or she is still alive in or outside the hospital  $\Pr(T_k^1 \geq u | T_k^1 \geq t, \mathcal{D}(t), X_k)$ . This does not take into account that discharged patients are no longer at risk of dying in the hospital after their time of discharge. So we developed extra routines to calculate the probability that a patient will die in the hospital given he or she has not yet been discharged  $\Pr(T_k^1 \geq u | T_k^1 \geq t, T_k^2 \geq t, \mathcal{D}(t), X_k)$ .

A possible extension of our joint model would be to specify separate longitudinal submodels for each of the 6 organ system scores. In this way the prognostic value of each score can be assessed, which might provide additional information to help decision-making. However, currently within the JM package, only one organ system score can be handled at a time and so it is impossible to account for likely associations between the different scores. Some applications of the joint modelling technique to analyze multiple longitudinal outcomes simultaneously have been discussed in the literature (Chi and Ibrahim 2006, Musoro et al., 2014, Hof et al., 2016).

We implemented the standard parameterization used in joint modelling which assumes that

the hazard at a given time point  $t$  depends on the expected SOFA score at the same time point  $t$ . Alternative parameterizations have been discussed in the literature (Rizopoulos 2012 and Rizopoulos et al., 2013 ). For instance, it can be assumed that the hazard depends on the rate of change of the SOFA scores via a slope term that is calculated using the available SOFA scores up till time point  $t$  or the area under the SOFA score trajectory up to  $t$ . In future work it will be interesting to examine how these different parameterizations may influence the predictive ability of the SOFA score.

Another popular approach for calculating dynamically-updated predictions of survival probabilities with time-dependent covariates is by landmarking (Van Houwelingen, 2007; Yingye and Heagerty, 2005; Van Houwelingen and Putter, 2012). This approach has also been applied to data with competing risks (Nicolai et al., 2013, 2012; Cortese and Andersen, 2010). Here, at every landmark time point  $t_s$ , a survival model is fitted using data of individuals who are at risk at landmark  $s$ . This model includes the marker history until  $t_s$  as a time-fixed covariate. Survival probabilities can be computed at the desired prediction horizon  $t_{hor}$  conditional on updated information at  $t_s$ . This approach can be easily extended to accommodate several markers. But in a setting that involves only a single longitudinal marker, it was observed that there was a substantial gain in predictive performance from using a joint model instead of landmarking for a situation in which the joint model was correct (Rizopoulos et al., 2013). In conclusion, the joint modelling framework offer clinicians with a valuable additional predictive tool to better understand and quantify the severity of illness and risk of death amongst IC patients.

## References

- Beyersmann, J. and M. Schumacher (2008). Time-dependent covariates in the proportional subdistribution hazards model for competing risks. *Biostatistics*, 4:765–76.
- Cabr, L., J. Mancebo, J. Solsona, P. Saura, I. Gich, L. Blanch, G. Carrasco, and M. Martn (2005). Multicenter study of the multiple organ dysfunction syndrome in intensive care units: the usefulness of sequential organ failure assessment scores in decision making. *Intensive Care Medicine*, 31:927–933.
- Chi, Y. and Ibrahim, G. (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62:432–45.

- Cortese, G. and P. K. Andersen (2010). Competing risks and time-dependent covariates. *Biometrical journal*, 52:138–58.
- Deslandes, E. and S. Chevret (2010). Joint modeling of multivariate longitudinal data and the dropout process in a competing risk setting: application to ICU data. *BMC Medical Research Methodology*, 10(1):69.
- Elashoff, R. M., G. Li, and N. Li (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics*, 64:762–71.
- Faucett, C. L. and D. C. Thomas (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in medicine*, 15:1663–85.
- Fine, J. P. and R. J. Gray (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 446:496–509.
- Geskus, R. B. (2011). Cause-Specific Cumulative Incidence Estimation and the Fine and Gray Model Under Both Left Truncation and Right Censoring. *Biometrics*, 67:39–49.
- Geskus, R. B. (2014). Which individuals make dropout informative? *Stat. Meth. in Med. Research*, 23:91–106.
- Gueorguieva, R., R. Rosenheck, and H. Lin (2012). Joint modelling of longitudinal outcome and interval-censored competing risk dropout in a schizophrenia clinical trial. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, 175:417–433.
- Ho, K., K. Lee, T. Williams, J. Finn, M. Knuiman, and S. Webb (2007). Comparison of acute physiology and chronic health evaluation (APACHE) II score with organ failure scores to predict hospital mortality. *Anaesthesia*, 62:466–473.
- Hof M., Musoro J., Geskus R., Struijk G., ten Berge I. Zwinderman A. (2016). Simulated maximum likelihood estimation in joint models for multiple longitudinal markers and recurrent events of multiple types, in the presence of a terminal event *Journal of Applied Statistics*, 1–22.
- Kajdacsy-Balla, A. A., F. Andrade, R. Moreno, A. Artigas, F. Cantraine, and J. Vincent (2005). Use of the sequential organ failure assessment score as a severity score. *Intensive Care Medicine*, 31:243–249.

- Khwannimit, B. (2007). A comparison of three organ dysfunction scores: MODS, SOFA and LOD for predicting ICU mortality in critically ill patients. *J Med Assoc Thai.*, 90:1074–1081.
- Li, N., R. M. Elashoff, G. Li and J. Saver (2010). Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial. *Statistics in Medicine*, 29:546–57.
- Minne, L., A. Abu-Hanna, and E. de Jonge (2008). Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review. *Critical Care*, 12:R161.
- Moreno, R., J.-L. Vincent, R. Matos, A. Mendona, F. Cantraine, L. Thijs, J. Takala, C. Sprung, M. Antonelli, H. Bruining, and S. Willatts (1999). The use of maximum sofa score to quantify organ dysfunction/failure in intensive care. results of a prospective, multicentre study. *Intensive Care Medicine*, 25:686–696.
- Musoro, J. Z., R. B. Geskus, and A. H. Zwinderman (2015). A joint model for repeated events of different types and multiple longitudinal outcomes with application to a follow-up study of patients after kidney transplant. *Biometrical Journal*, 57(2):185–200.
- Nicolaie, M. A., H. Van Houwelingen, T. M. de Witte, and H. Putter (2012). Dynamic prediction by landmarking in competing risks. *Statistics in Medicine*, 12:2031–47.
- Nicolaie, M. A., H. Van Houwelingen, T. M. de Witte, and H. Putter (2013). Dynamic pseudo-observations: A robust approach to dynamic prediction in competing risks. *Biometrics*, 4:1043–52.
- Njagi, E. N., D. Rizopoulos, G. Molenberghs, P. Dendale, and K. Willekens (2013). A joint survival-longitudinal modelling approach for the dynamic prediction of rehospitalization in telemonitored chronic heart failure patients. *Statistical Modelling*, 13:179–198.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing* <http://www.R-project.org/>. R Foundation for Statistical Computing, Vienna, Austria.
- Rizopoulos, D. (2010). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9):1–33.
- Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67:819–29.

- Rizopoulos, D. (2012). Joint Models for Longitudinal and Time-to-Event Data: With Applications in R *Boca Raton: Chapman and Hall/CRC*.
- Rizopoulos, D., M. Murawska, E. Andrinopoulou, G. Molenberghs, J. J. M. Takkenberg, and E. Lesaffre (2013). Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *arXiv:1306.6479v1*, pages 1–34.
- Toma, T., A. Abu-Hanna, and R. Bosman (2007). Discovery and inclusion of SOFA score episodes in mortality prediction. *Journal of Biomedical Informatics*, 40:649–660.
- Toma, T., A. Abu-Hanna, and R. Bosman (2008). Discovery and integration of univariate patterns from daily individual organ-failure scores for intensive care mortality prediction. *Artif Intell Med.*, 43:47–60.
- Tsiatis, A. A. and M. Davidian (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14:809–834.
- Van Houwelingen, H. (2007). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*, 34:70–85.
- Van Houwelingen, H. and H. Putter (2012). *Dynamic prediction in clinical survival analysis*. Chapman & Hall, Boca Raton.
- Vincent, L. and L. Ferreira (2000). Evaluation of organ failure: we are making progress. *Intensive Care Medicine*, 26:1023–4.
- Wulfsohn, M. S. and A. A. Tsiatis (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53:330–339.
- Xu, J. and S. Zeger (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50:375–387.
- Yingye, Z. and P. J. Heagerty (2005). Partly conditional survival models for longitudinal data. *Biometrics*, 61:379–391.

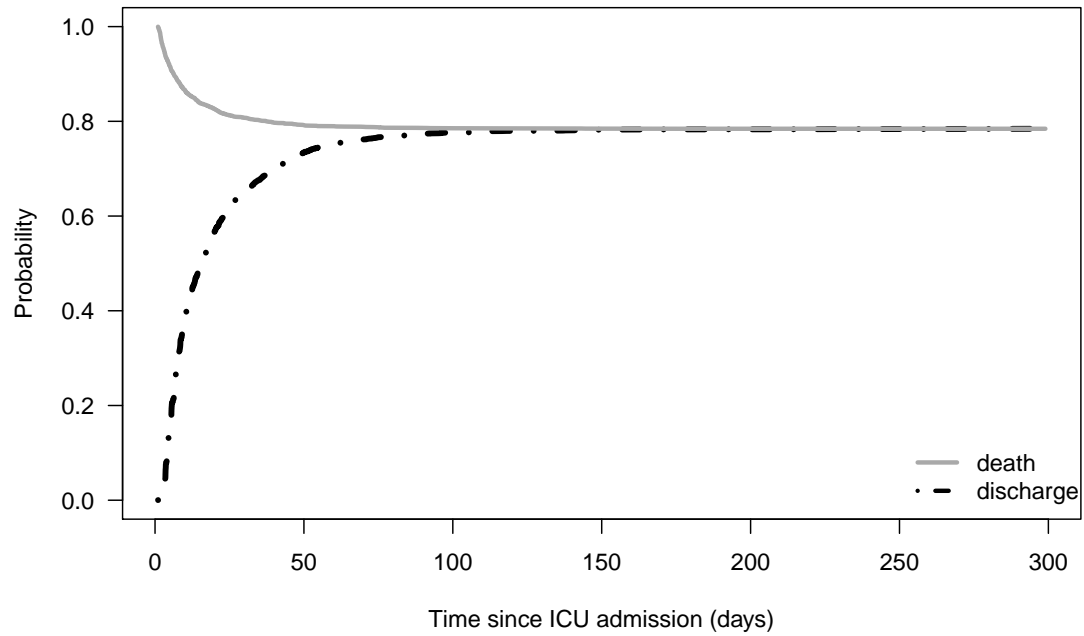


Figure 1: The cumulative incidence function for death and discharge. We use the alternate display format where the plot of the cause-specific probability of dying starts at 1 and decreases, while that for discharge starts at 0 and increases.



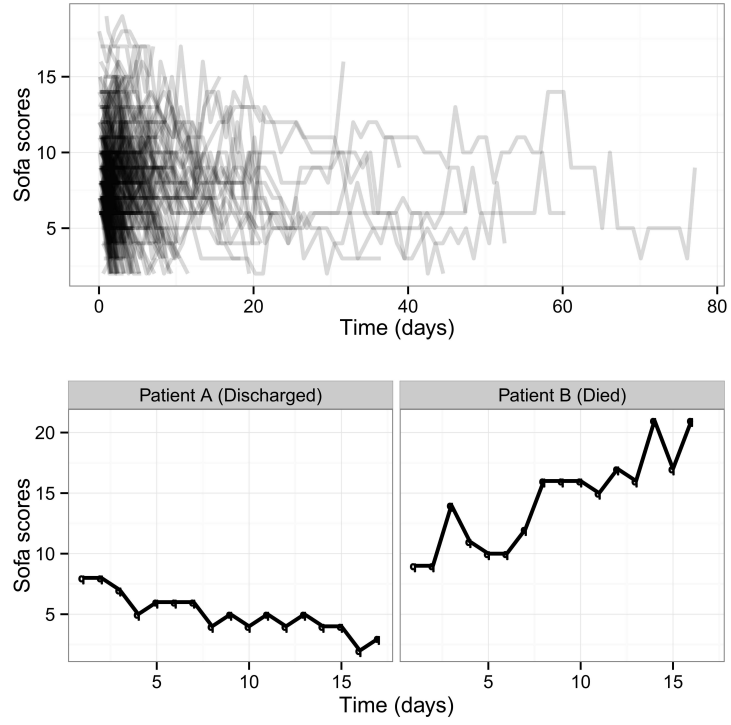


Figure 2: Individual SOFA score profiles for 500 randomly selected subjects (upper panel). The lower panel shows the trajectory of Patient A who was discharged from the hospital and Patient B who died in the ICU.

Table 1: Parameter estimates from the joint model with death as event.

	<i>Coefficients</i>	<i>standard error</i>
Longitudinal submodel		
<i>Intercept</i>	4.6754	0.0940
$\beta_{time}$	-0.3504	0.0016
$\beta_{SAPS}$	0.1023	0.0019
Survival submodel		
$\beta_{SAPS}$	0.1053	0.0014
$\beta_{SOFA}$	0.3679	0.0092

All parameters had p value <0.0001

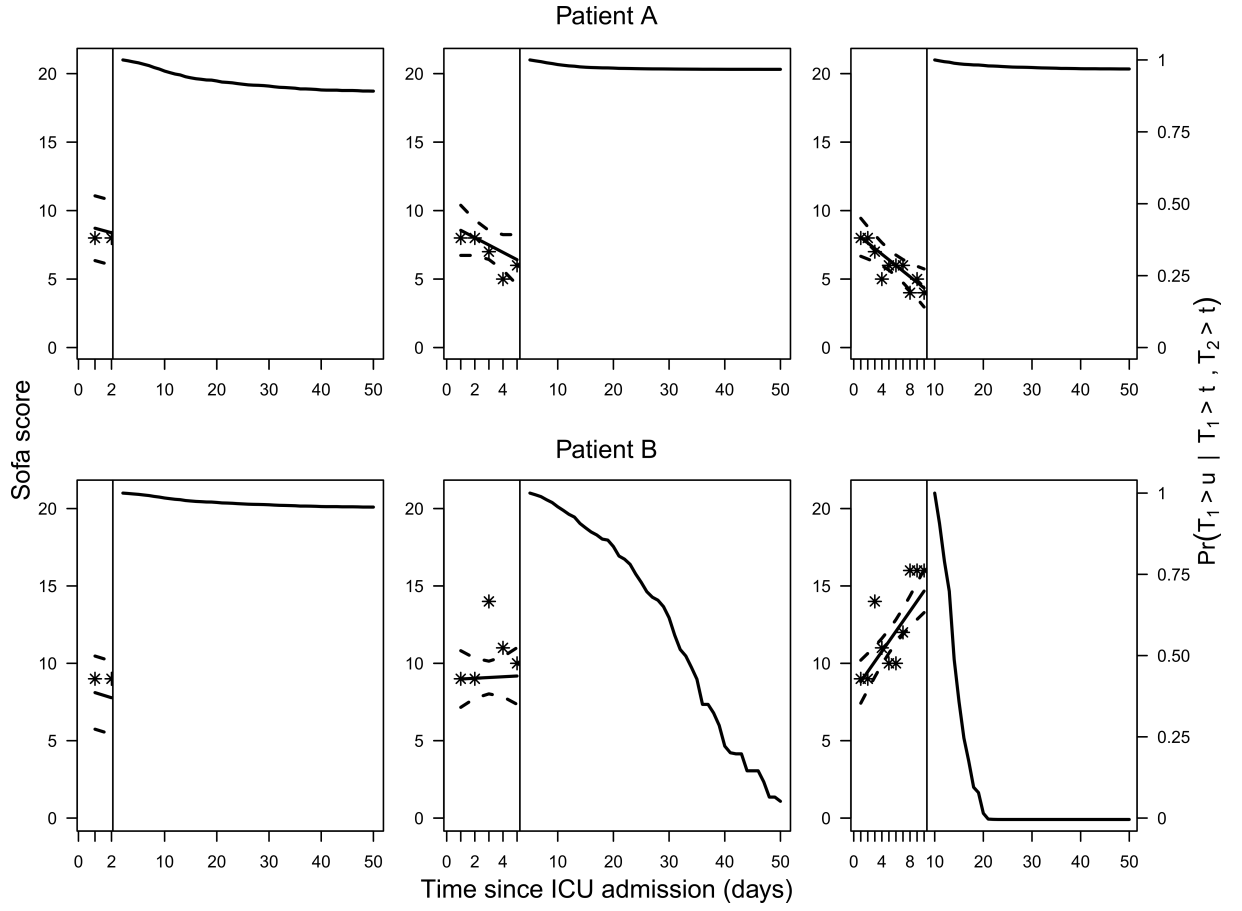


Figure 3: Dynamic survival probabilities for patients A (upper panel) and patients B (lower panel) based on SOFA scores available up till days 2, 5 and 10 respectively. Notice that we use different time scales for the SOFA scores (left) and the survival probability (right).

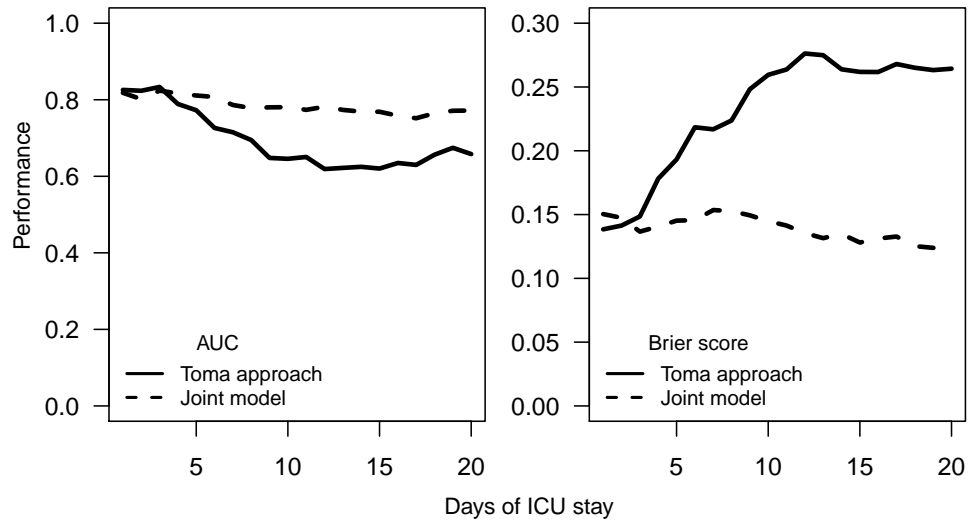


Figure 4: Test performances, AUC's (left panel) and Brier scores (right panel), from Toma's approach (solid lines) and the joint modelling approach (broken lines).