Statistical Modelling 2016; **16(5)**: 409–428

Time-dependent ROC methodology to evaluate the predictive accuracy of semiparametric multi-state models in the presence of competing risks: An application to peritoneal dialysis programme

Laetitia Teixeira[1,2], Carmen Cadarso-Suárez[3], Anabela Rodrigues[1,4] and Denisa Mendonça[1,2]
[1]Institute of Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal
[2]ISPUP-EPIUnit, University of Porto, Porto, Portugal
[3]Unit of Biostatistics, Department of Statistics and Operations Research, School of Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain
[4]Nephrology Department, CHP-Hospital Geral de Santo António, Porto, Portugal

Abstract: The evaluation of peritoneal dialysis (PD) programmes requires the use of statistical methods that suit the complexity of such programmes. Multi-state regression models taking competing risks into account are a good example of suitable approaches. In this work, multi-state structured additive regression (STAR) models combined with penalized splines (P-splines) are proposed to evaluate peritoneal dialysis programmes. These models are very flexible since they may consider smooth estimates of baseline transition intensities and the inclusion of time-varying and smooth covariate effects at each transition. A key issue in survival analysis is the quantification of the time-dependent predictive accuracy of a given regression model, which is typically assessed using receiver operating characteristic (ROC)-based methodologies. The main objective of the present study is to adapt the concept of time-dependent ROC curve, and their corresponding area under the curve (AUC), to a multi-state competing risks framework. All statistical methodologies discussed in this work were applied to PD survival data. Using a multi-state competing risks framework, this study explored the effects of major clinical covariates on survival such as age, sex, diabetes and previous renal

of STAR models combined with time-dependent ROC curves revealed important conclusions not previously reported in the nephrology literature when using standard statistical methodologies. For practical application, all the statistical methods proposed in this article were implemented in `R` and we wrote and made available a script named as `NestedCompRisks`.

Key words: time-dependent ROC curve; competing risks; multi-state models; peritoneal dialysis; STAR model; survival analysis

Received April 2014; **revised** June 2016; **accepted** June 2016

Address for correspondence: Laetitia Teixeira, Institute of Biomedical Sciences Abel Salazar, University of Porto, Rua de Jorge Viterbo Ferreira, 228 4050-313 Porto, Portugal.
E-mail: lcteixeira@icbas.up.pt

10.1177/1471082X16658731

1 Introduction

End-stage renal disease is becoming a major public health problem with a growing number of patients in need of replacement therapies, namely renal transplant, haemodialysis and peritoneal dialysis (Levey *et al.*, 2007). Similarly to patients suffering from other chronic diseases, renal patients under peritoneal dialysis (PD) are periodically monitored and relevant clinical data such as the presence/absence of comorbidities and the occurrence of peritonitis or hospitalizations are often collected (Davenport, 2009; Martins *et al.*, 2013; Rocha *et al.*, 2012; Teixeira *et al.*, 2013). The competing events—death, transfer to haemodialysis and renal transplantation—are the main causes of dropout of patients from a PD programme, leading to the definition of the following indicators to evaluate such PD programmes: patient survival (considering death as the endpoint of interest) and technique survival (considering transfer to haemodialysis as the endpoint of interest).

Given the complexity of the disease evolution and the available relevant clinical information, the evaluation of a PD programme requires, from a statistical point of view, a suitable approach, such as a multi-state approach taking competing risks into account. Several competing risks definitions have been proposed and the definition proposed by Gooley *et al.* (1999) was adopted in the present study: a competing risk is '... an event whose occurrence either precludes the occurrence of another event under examination or fundamentally alters the probability of occurrence of this other event'. Multi-state modelling is an adequate alternative approach to the classical survival models and presents several advantages. One of these advantages is the possibility of evaluating how specific prognostic factors may influence different phases of the disease progression, which is usually ignored in classical approach of survival analysis (de Wreede *et al.*, 2011; Hougaard, 1999; Meira-Machado *et al.*, 2009). The multi-state approach allows a more detailed description of the patient trajectory, enabling a better biological knowledge about the disease/recovery process.

Although increasing importance has been given to the competing risk and multi-state approaches in clinical and epidemiological research (see, e.g., Andersen and Keiding, 2012), the majority of the literature on PD depicts only classic survival methods being used (Perl *et al.*, 2012; Yang *et al.*, 2013). Therefore, further developments are needed to account for the complexity of this clinical setting. In recent years, a variety of flexible survival regression methods based on various statistical models have been proposed when a set of covariates is present: the additive hazards approach of Aalen (Martinussen and Scheike, 2010), Additive Cox regression models (Hastie and Tibshirani, 1990a,b), and structured additive regression (STAR) models to the analysis of survival data (Brezger and Lang, 2006; Hennerfeind *et al.*, 2006; Kneib and Fahrmeir, 2007; Kneib and Hennerfeind, 2008). In this study, we used the general semiparametric class of multi-state models based on the STAR, proposed by Kneib and Hennerfeind (2008), using penalized splines (P-splines). These models allow a flexible modelling of baseline transition intensities in terms of P-splines and the inclusion of parametric, time-varying and non-parametric covariate effects.

The evaluation of the predictive accuracy of a survival model is one of the most important considerations in the development of a prediction model (Chen

et al., 2012). In a multi-state framework, this aspect is crucial since it allows the identification of different prognostic factors for each transition that composes a multi-state model. One of the most popular methods to evaluate the prognostic ability of a survival regression model is the analysis of classification measures such as sensitivity and specificity through receiver operating characteristic (ROC) curves (Pepe, 2004; Swets and Pickett, 1982; Zhou *et al.*, 2011), taking a model score as the diagnostic marker. In the last years, there has been an increasing interest in extending the standard binary classification accuracy measures—like sensitivity, specificity, true-positive and false-positive—to the survival context (see, e.g., Etzioni *et al.*, 1999; Heagerty *et al.*, 2000; Heagerty and Zheng, 2005; Pepe *et al.*, 2008). These proposed methodologies lead to time-dependent definitions, resulting in the so-called time-dependent ROC curves. Two different extensions for classification measures are proposed in the literature, both corresponding to the consideration of *cumulative* (prevalent) cases recruited over a fixed period (Heagerty *et al.*, 2000) or alternatively, to *incident* cases that are observed for any selected time t (Etzioni *et al.*, 1999; Heagerty and Zheng, 2005). To evaluate the predictive accuracy of the multi-state model, we will follow the definition of Heagerty *et al.* (2000) for the time-dependent ROC curve, adapting their methodology to the multi-state framework in the presence of competing risks. This way, a different ROC curve is obtained for each transition, which allows us to compare the prognostic capability of the model for each transition.

The main objectives of this study are: (a) to adapt the definition of the time-dependent ROC curve in order to evaluate the predictive accuracy of STAR models in a multi-state competing risks framework (Kneib and Hennerfeind, 2008) and (b) to show the relevance of this approach in the analysis of a PD programme.

The article is organized as follows: After this introduction, in Section 2, we describe a multi-state model as a succession of nested competing risks models. We also define the STAR model and flexible hazard ratio (HR) curves for continuous covariates. Time-dependent ROC curves to assess the predictive accuracy of the STAR models are described in Section 3. The application of the proposed methodologies in PD patients is discussed in Section 4. Finally, a general discussion is presented in Section 5.

2 Semiparametric competing risks multi-state models

A multi-state process with several absorbing states is a stochastic process $(X(t), t \in T)$ in continuous time with a finite space of states $S = \{1, \ldots, K\}$ where T represents survival time. This model is completely characterized by the transition intensity between two states h_1 and h_2 ($h_1, h_2 = 1, \cdots, K$). The transition will be designated from now on by h.

The transition intensity (or transition hazard) between the states h_1 and h_2, λ_h, is given by the cause-specific hazard function:

$$\lambda_h(t) = \lim_{\triangle t \to 0} \frac{P(X(t + \triangle t) = h_2 | X(t) = h_1)}{\triangle t} \tag{2.1}$$

A basic competing risks model is a model composed by $k + 1$ states, that is one initial state and k absorbing states that represent k endpoints. All transitions (state changes) start in the state 0 (initial state) until one of the k absorbing states (Andersen and Keiding, 2012; de Wreede *et al.*, 2011; Meira-Machado *et al.*, 2009; Putter *et al.*, 2007). These competing risks models are also characterized by the fact that only the initial state is not an absorbing state, that is, absence of transient states (Hougaard, 1999). However, in some situations the clinical trajectories of the patients are complex and cannot be explained using a basic competing risks model. In these situations, it is necessary to use more complex models, such as a multi-state model composed by one (or more) initial states, one (or more) transient states and two (or more) absorbing states. This situation is observed in the context of PD, where patients may experience a non-absorbing event (e.g., peritonitis) before the observation of an absorbing event (death, transfer to haemodialysis or renal transplantation). For a multi-state model with several absorbing states—that is, in the presence of competing risks—Beyersmann *et al.* (2012) proposed an algorithm that describes a multi-state model through a succession of nested competing risks models. The main advantage of this approach is the use of the methodology of competing risks in multi-state models.

As already referred to in the Introduction, it is quite common to include a set of covariates in the multi-state model (2.1). Quite often, it is assumed that the effects of continuous covariates have a linear functional form in all transitions. However, this assumption may not be appropriate for real data applications, in which the effects of continuous covariates are generally unknown (Cadarso-Suárez *et al.*, 2010; Meira-Machado *et al.*, 2013). Based on the STAR models, Kneib and Hennerfeind (2008) proposed a general semiparametric class of multi-state models.

The additive regression model can be written as the exponential of the additive combination of $J + 1$ components:

$$\lambda_h(t|\mathbf{z}) = \exp(\eta_h(\mathbf{z})) = \exp(g_{h0}(t) + \sum_{j=1}^{J} f_{hj}(\mathbf{x})) \tag{2.2}$$

where η_h is the structured additive predictor and $\mathbf{z} = (t, x_1, \ldots, x_J)'$ is the vector containing both, the observed time and the vector of J covariates of different types (continuous and categorical covariates).

In equation (2.2), $g_{0h}(t)$ represents the log-baseline hazard rate ($g_{h0}(t) = \log(\lambda_{h0}(t))$) and f_{hj} are generic representations of different types of covariate effects: linear effects of a continuous covariate x, $f_h(x) = x\beta_h$; categorical effects of a dummy coding v of a categorical covariate x, $f_h(x) = v'\beta_h$; non-parametric, smooth effects of a continuous covariate x, $f_h(x) = s_h(x)$; time-varying effects of a categorical or continuous covariate x, $f_h(x) = xs_h(t)$.

All the flexible (non-parametric) effects, including the log-baseline hazard, are all modelled using P-splines (Eilers and Marx, 1996). The general idea is to approximate the functions g_{h0} and f_{hj} as linear combinations of basis splines (B-splines) basis functions:

$$f_{hj}(x) = \sum_{q=1}^{d_k} \beta_{hq} B_{hq}(x) \tag{2.3}$$

where vector $\beta_{hj} = (\beta_{h1}, \ldots, \beta_{hd_k})$ is the vector of unknown regression coefficients corresponding to the B-splines basis of degree a and defined over a grid of k knots lying on the domain of x, with $d_k = a + k - 1$. Each predictor component can be expressed as the product of an appropriate design matrix $\mathbf{X}_{hj}$ composed of basis function evaluations and the vector β_{hj} of regression coefficients. Moreover, a penalty term is added to control the level of smoothness by penalizing wiggly functions when estimating β_{hj}. The most commonly used penalization term is based on the integral of the second derivative of the smooth functions, f_{hj}. We consider a discrete approximation of it given by $pen(f_{hj}) = \lambda_{hj}\beta'_{hj}\mathbf{K}_{hj}\beta_{hj}$, where the matrix $\mathbf{K}_{hj}$ is a positive semi-definite matrix, that can be written as $\mathbf{K}_{hj} = \mathbf{D}'_{hj}\mathbf{D}_{hj}$ (with $\mathbf{D}_{hj}$ the second order difference matrix of neighbouring), and $\lambda_{hj} \geq 0$ a smoothing parameter (Eilers and Marx, 1996).

The estimation of the regression effects is based on the penalized log-likelihood derived from the representation of the smooth effects in terms of P-splines, that is, each model is fitted by maximizing:

$$l_{\mathsf{pen}}(\beta_h) = l(\beta_h) - \sum_{j=1}^{p} \lambda_{hj}\beta'_{hj}\mathbf{K}_{hj}\beta_{hj} \tag{2.4}$$

Estimation can be based on a unified Bayesian formulation that incorporates penalized splines and random effects into one general framework. Different approaches can be used for Bayesian inference: using Markov Chain Monte Carlo simulation techniques or using mixed model representations of STAR models for empirical Bayesian inference. In this work, we use empirical Bayes inference, where the variance parameters τ^2 are treated as fixed unknown constants to be estimated from their marginal posterior (Kneib and Hennerfeind, 2008). The smoothing parameters are considered as variance components corresponding to the vector of regression coefficients. This methodology allows the simultaneous estimation of the regression coefficients and of the smoothing parameters corresponding to each unknown function g_{h0} or f_{hj} using restricted maximum likelihood (REML) estimation. More details can be found in the study of Kneib and Hennerfeind (2008). The implementation of this approach is available in the software package *BayesX* (Belitz *et al.*, 2012).

In order to obtain interpretable results achieved from the application of STAR models, assuming the semiparametric multi-state model in (2.2), flexible HR curves can be obtained. These HR curves describe the relationship between each of the continuous covariates and the outcome in each transition, when a specific value of the covariate is taken as reference (see Cadarso-Suárez *et al.*, 2010). In the

multi-state regression framework, it is recommended (Cadarso-Suárez *et al.*, 2010; Meira-Machado *et al.*, 2013) to take as the reference a common clinical reference value or some value related to clinical normality. These HR curves—along with their corresponding pointwise confidence bands—can be obtained using the `R`-based package `smoothHR` (Cadarso-Suárez *et al.*, 2010; Meira-Machado *et al.*, 2013).

3 Evaluating the predictive accuracy of the multi-state model: Time-dependent ROC curves

In survival analysis framework, one of the major interests is to predict the outcome based on various factors. Then, measures of the predictive accuracy of survival regression models need to be considered. These measures quantify the extent to which covariates determine an individual outcome (Schemper, 2003). ROC analysis is an effective method of evaluating the quality or performance of a model. ROC curves plot the sensitivity (true positive rate) versus the 1-specificity (false positive rate).

In the last few years, several methods have been proposed to characterize the predictive accuracy of a classical survival regression model when the outcome of interest is a censored survival time, that is, when the patient's status changes over time (see, e.g., Etzioni *et al.*, 1999; Heagerty *et al.*, 2000; Heagerty and Zheng, 2005; Pepe *et al.*, 2008). Time-dependent ROC curves can be used to summarize the accuracy of a classical survival model, offering an alternative to the use of, for example, the proportion of variation explained for censored data models. In this context, several extensions of time-dependent sensitivity and specificity were proposed (Heagerty and Zheng, 2005). In competing risks framework, we need to take into account these concepts (Blanche *et al.* 2013; Heagerty, 2010; Saha and Zheng *et al.*, 2012).

Let T_i denote failure time and C_i the censoring time. $Z_i^* = \min(T_i, C_i)$ represents the follow-up time and δ_i a censoring indicator, $\delta_i = \{0, 1\}$. In this work, the cumulative sensitivity and dynamic specificity definitions used in the setting of classical survival were adapted to our multi-state modelling approach based on the definitions proposed by Heagerty *et al.* (2000). Specifically, by considering a scalar marker value Z_h that represents $\sum_{j=1}^{J} f_{hj}(\mathbf{z})$ from the STAR model for the transition h as represented in expression (2.2), and survival time T represented through the counting process $N_h^*(t) = 1(T \leq t)$, cumulative sensitivity and dynamic specificity for the transition h are defined as:

$$\begin{aligned} &\text{sensitivity}_h(c, t) : P(Z_h > c | T \leq t, \delta = h) = P\{Z_h > c | N_h^*(t) = 1\} \\ &\text{specificity}_h(c, t) : P(Z_h \leq c | T > t) = P\{Z_h \leq c | N_h^*(t) = 0\} \end{aligned} \tag{3.1}$$

Using these definitions, each person is classified as either a case or a control on the basis of the transition h at time t, defining the corresponding ROC curve for any time t, ROC(t) (Heagerty *et al.*, 2000; Heagerty and Zheng, 2005). Considering the definitions of cumulative sensitivity and dynamic specificity adapted to a multi-state

framework shown in equations (3.1), ROC curves are defined as:

$$\mathrm{ROC}_t(p) = \mathrm{TP}_t\{[\mathrm{FP}_t]^{-1}(p)\} \tag{3.2}$$

where $\mathrm{TP}_t(c) = P\{Z_h > c | N_h^*(t) = 1\}$, $\mathrm{FP}_t(c) = P\{Z_h > c | N_h^*(t) = 0\}$ and $[\mathrm{FP}_t(p)]^{-1} = \inf_c\{c : \mathrm{FP}_t(c) \leq p\}$. These ROC curves measure the predictive accuracy of the scalar marker value Z_h to distinguish between subjects who experience the particular transition by time t and those who do not experience such a transition (Saha and Heagerty, 2010).

In the presence of censored survival times, Heagerty *et al.* (2000) also developed a non-parametric estimator of sensitivity and specificity based on the nearest-neighbour bivariate survival estimator proposed by Akritas (1994). Such estimator can be adapted to the context of multi-state models. Explicitly, considering the bivariate distribution function $S_h(c, t)$ at transition h:

$$S_h(c, t) = P(Z_h > c, T > t) = \int_c^{\infty} S_h(t | Z_h = s) dF_{Z_h}(s) \tag{3.3}$$

where $F_{Z_h}(s)$ is the distribution function for Z_h, the nearest-neighbour estimator (NNE) is then given by:

$$\hat{S}_{h\gamma_n}(c, t) = \frac{1}{n} \sum_i \hat{S}_h \gamma_n(t | Z_h = Z_{hi}) \mathbf{1}(Z_{hi} > c) \tag{3.4}$$

where $\hat{S}_{h\gamma_n}(t | Z_h = Z_{hi})$ is an estimator of the conditional survival function with parameter γ_n, required to obtain a smooth estimate of $S(t | Z_h = Z_{hi})$. The weighted KM estimator is defined as:

$$\hat{S}_{h\gamma_n}(t | Z_h = Z_{hi}) = \prod_{s \in \tau_n, s \leq t} \left\{ 1 - \frac{\sum_j K_{\gamma_n}(Z_{hj}, Z_{hi}) \mathbf{1}(Z_j^* = s) \delta_j}{\sum_j K_{\gamma_n}(Z_{hj}, Z_{hi}) \mathbf{1}(Z_j^* \geq s)} \right\} \tag{3.5}$$

where $K_{\gamma_n}(Z_{hj}, Z_{hi})$ is a kernel function that depends on the smoothing parameter γ_n, while τ_n are the unique values of Z_i for observed events, $\delta_i = 1$. A 0/1 nearest-neighbour kernel was used, $K_{\gamma_n}(Z_{hj}, Z_{hi}) = \mathbf{1}\{-\gamma_n < \hat{F}_{Z_h}(Z_{hi}) - \hat{F}_{Z_h}(Z_{hj}) < \gamma_n\}$, where $2\gamma_n \in (0, 1)$ represents the percentage of observations that is included in each neighbourhood (Akritas, 1994).

The estimates of sensitivity and specificity here proposed for the context of multi-state models are given by:

$$\hat{P}_{h\gamma_n}\left\{Z_h > c | N_h^*(t) = 1\right\} = \frac{\left\{1 - \hat{F}_{Z_h}(c)\right\} - \hat{S}_{h\gamma_n}(c, t)}{1 - \hat{S}_{h\gamma_n}(t)}$$
$$\hat{P}_{h\gamma_n}\left\{Z_h \leq c | N_h^*(t) = 0\right\} = 1 - \frac{\hat{S}_{h\gamma_n}(c, t)}{\hat{S}_{h\gamma_n}(t)} \tag{3.6}$$

where $\hat{S}_h\gamma_n(t) = \hat{S}_h\gamma_n(-\infty, t)$.

Considering these two estimators in (3.6), the estimated ROC_t curve is given by:

$$\hat{\mathrm{ROC}}_t(p) = \hat{\mathrm{TP}}_t\left\{[\hat{\mathrm{FP}}_t]^{-1}(p)\right\} \tag{3.7}$$

Several summary indices are associated with the time-dependent ROC curve (3.2). One of the most popular measures is the time-varying area under the ROC curve (AUC(t)) that measures, at each time t, the probability that the marker value for a randomly selected case exceeds the marker value for a randomly selected control (Zhou *et al.*, 2011). Explicitly, the AUC(t) is given by $\mathrm{AUC}(t) = \int_0^1 \mathrm{ROC}_t(p)dp$. Using $(\hat{\mathrm{ROC}})_t(p)$ in (3.7), AUC(t) can be estimated by $\hat{\mathrm{AUC}}(t) = \int_0^1 \hat{\mathrm{ROC}}_t(p)dp$.

For each t, a 95% confidence interval (CI) for the AUC(t) can be obtained using resampling methods dealing with censored data (Davison and Hinkley, 1997; Efron, 1981). In this work, Cox model-based bootstrap procedures (see Davison and Hinkley (1997), pp. 351–58) were used as the resampling technique. The simplest type of resampling was used which resamples with replacement from the observations, using the `R`-based package `censboot`.

4 Analysis of the peritoneal dialysis programme

Patients with chronic kidney disease in the PD programme are periodically monitored during the follow-up time. As recovery of renal function is very rare, patients can transit only to one of the following absorbing states: death, transfer to haemodialysis or renal transplantation. To analyze the clinical trajectory of these patients, it is necessary to use a competing risks approach.

We consider a model with the structure schematically presented in Figure 1(a). The initial state (state 0) corresponds to the moment when the patient enters the PD programme. After the beginning of the programme, the patients can transit to one of the three absorbing events. The model also includes a transient state, which represents the occurrence of at least one episode of peritonitis during the PD programme.

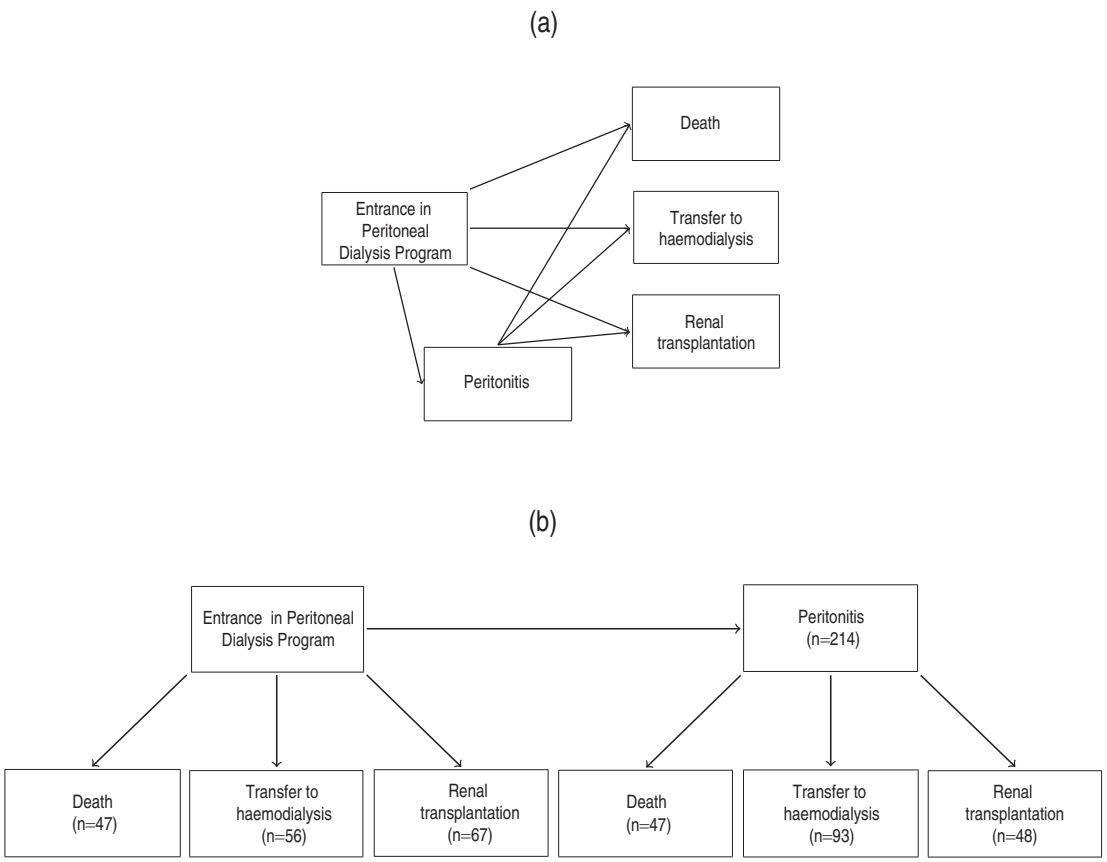


Figure 1 (a) Multi-state model with a transient state (peritonitis) for PD patients; (b) Representation of nested competing risks models.

In order to apply the techniques used in the competing risks approach, the model presented in Figure 1(a) was decomposed in nested competing risks models as shown in Figure 1(b), using the approach proposed by Beyersmann *et al.* (2012). Specifically, this model is composed of an initial state (entrance in the PD programme), a transient state (occurrence of at least one episode of peritonitis) and several absorbing states (death, transfer to haemodialysis and renal transplant, without and with a peritonitis episode).

This study considered all patients (n = 427) with chronic kidney disease included in the PD programme of the Peritoneal Dialysis Unit of the Nephrology Department, Hospital Geral de Santo António—Centro Hospital do Porto (Portugal) between January 1980 and July 2011. The patients under study who experienced an event are distributed according to the states as represented in Figure 1(b). In addition, socio-demographic and clinical characteristics were considered: age, gender, diabetes and previous renal replacement therapy (PRRT) (yes/no, i.e., some patients could

previously use another renal replacement therapy before PD, such as haemodialysis or renal transplantation).

STAR models were fitted using the *BayesX* software (Belitz *et al.*, 2012). All other analyses were performed with `R` (RCoreTeam, 2013) software. We offer a script which we named as `NestedCompRisks` to perform a similar analysis on any data set, requiring only the specification of the number of transitions of the data set at hand and considering some additional features of the data: to create a dummy variable for each transition and a state variable indicating the state from which the transition begins. This script is composed of the following three functions: (a) `Expl_NCR`: to perform the exploratory analysis, including, for example, Nelson–Aalen and Aalen–Johansen estimators; (b) `Mult_NCR`: to give results obtained from fitting semiparametric multi-state models in the presence of competing risks. The output includes flexible hazard rate curves and their confidence bands and (c) `AUC_NCR`: to calculate the time-dependent predictive accuracy of the fitted model through time-dependent ROC curves, and the time-varying AUCs, together with their corresponding bootstrap confidence intervals. The script can be obtained from the first author.

4.1 Structured additive multi-state modelling

With the purpose of obtaining a valid description of the trajectory of PD patients according to some patients' characteristics, a STAR model was implemented, considering the model described in Section 2.2, expression (2.2).

Based on the model structure represented in Figure 1, a STAR model using the *BayesX* software was considered:

$$\eta_b(\mathbf{z}) = g_{0b}(t) + s_b(age) + f_{b1}(gender) + f_{b2}(diabetes) + f_{b3}(PRRT) \qquad (4.1)$$

In our analyses, we considered B-spline of degree 3 and a grid of 20 equidistant knots (ensuring enough flexibility for the time-varying functions), taking the median as reference of the follow-up time for each transition, and a value of 55 years as the reference for the covariate age (Kotsanas *et al.*, 2007).

Parametric effects of the covariates gender, diabetes and PRRT are presented in Table 1. By analyzing the table, we find that diabetes is a significant predictor for the transitions Entrance → Death and Peritonitis → Death. Patients with diabetes, without or with experiencing a peritonitis episode, have a higher risk of death, compared to those patients without diabetes (HR = 2.42, 95% CI 1.29 – 4.51 and HR = 2.41, 95% CI 1.27 – 4.57, respectively). Previous renal replacement therapy is a significant predictor for the transitions Entrance → Death and Entrance → Peritonitis. Patients with a previous renal replacement therapy have a higher risk of death without peritonitis (HR = 1.93, 95% CI 1.05 – 3.53) and a higher risk of peritonitis (HR = 1.33, 95% CI 1.01 – 1.76) when compared with those patients without a previous renal replacement therapy.

To better understand the effects of age at each transition, we used HR curves and their confidence intervals and graphical results are shown in Figure 2. Patients older than 55 years have a significantly higher risk of death (without and with peritonitis)

and a significantly lower risk of transplant (without and with peritonitis) than those with 55 years (reference value). There is no evidence of statistical differences in the risk of transfer to haemodialysis (without and with peritonitis) for patients older than 55 years when compared to those under 55 years (Figure 2). Considering flexible log-baseline effects, only the transitions Entrance → Death and Peritonitis → Death showed a statistically significance effect (Table 1), that is, the transition rates for all the other transitions remained constant along the follow-up period time.

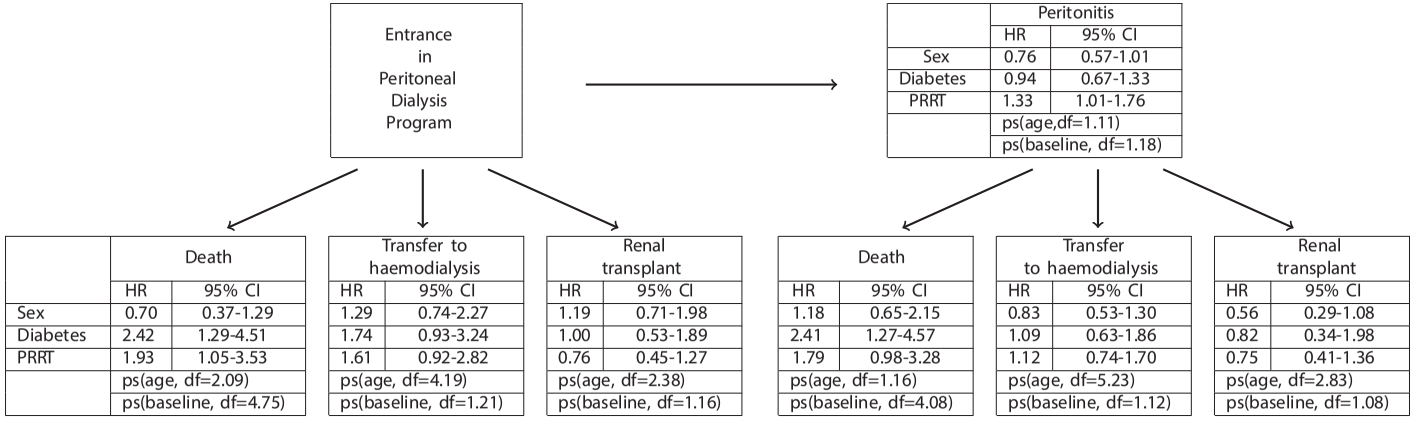
Given the relevance of presence of diabetes along follow-up, time-varying effects of diabetes were also tested and the graphical output of time-varying effects of diabetes for the transition entrance → death is shown in Figure 3. As it can be seen in this figure, diabetes has a statistically significant increasing time-varying effect at this transition, indicating that patients with diabetes present an increasing higher risk of death. For the other transitions, the risk remains constant over follow-up time.

4.2 Predictive accuracy of the STAR model—time-dependent ROC curves

In order to analyze the predictive accuracy of the model (4.1) time-dependent ROC curves were constructed at each transition (see Section 3). In all cases, a span of 0.05 was used in the NNE estimates of the bivariate survival function in (3.4). The time-varying AUCs for the ROC curves were then obtained, and their corresponding bootstrap 95% CIs were calculated, taking R = 1000 as the number of bootstrap replicates. Considering as long-term survival PD patients those who stay three or more years in the programme (Abraham *et al.*, 2010), Figure 4 presents the ROC curve for all transitions at t = 36 months. The AUC was higher for the transition Entrance → Death (AUC(36) = 0.795, 95% CI 0.715 – 0.878), followed by transition Peritonitis → Renal transplant (AUC(36) = 0.752, 95% CI 0.690 – 0.814) and Entrance → Renal transplant (AUC(36) = 0.709, 95% CI 0.634 – 0.792). For the transitions Entrance → Peritonitis and Entrance → Transfer to haemodialysis, the values of AUC(36) were very close to 0.50, thus indicating poor discriminative power for the corresponding models. Comparing the ROC curves for the three possible outcomes (death, transfer to haemodialysis and renal transplantation), differences were found considering the occurrence (or not) of a peritonitis episode. When at least one peritonitis episode was observed, renal transplant presented a larger AUC, indicating better discriminative power for the model. However, when no peritonitis was observed, death presented a larger AUC. In both situations, the outcome transfer to haemodialysis was the outcome with the smallest AUC associated, thus indicating poor discriminative power for the corresponding fitted models.

The discrimination ability of the model (4.1) was assessed by comparing the time-varying AUCs for the ROC curves that were generated for all transitions (Figure 5). First, notice that the accuracy of the model score remains good for the transitions with absorbing states death and renal transplant (Entrance → Death, Entrance → Renal transplant, Peritonitis → Death and Peritonitis → Renal transplant), with estimates of time-varying AUC between 0.6 and 0.8 over the follow-up period. Second, the discriminatory ability of the model remains constant for all transitions, with the exception of the transition Entrance → Death. For

Table 1 Estimates effects in each transition of the fitted multi-state model.



References: Sex-Female; Diabetes-No; PRRT-No

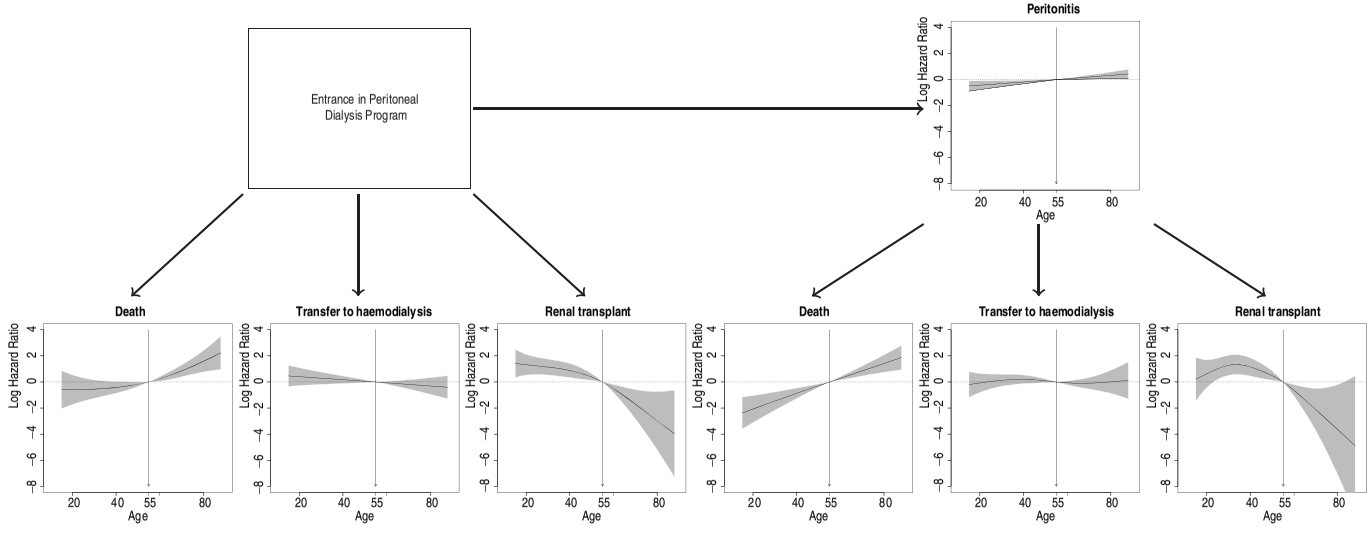


Figure 2 Adjusted smooth log hazard ratio estimates with 95% pointwise confidence bands for age (value 55 as reference) for STAR model.

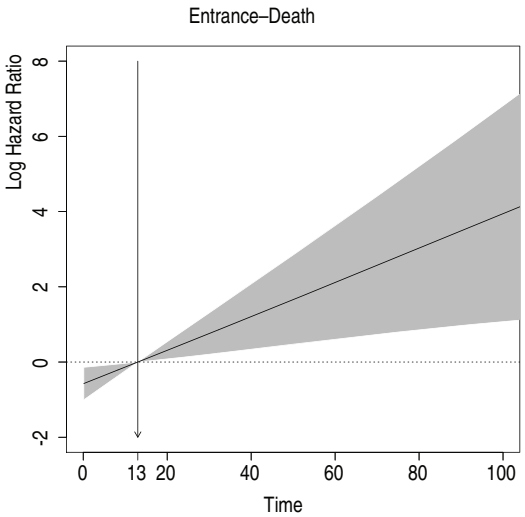


Figure 3 Time-varying effect of diabetes for transition Entrance → Death with reference value 13 months (HR(male) = 0.70, 95% CI 0.37 – 1.29; HR(PRRT) = 1.93, 95% CI 1.05 – 3.53; ps(age, df = 2.09); ps(baseline, df = 4.75); diabetes*ps(baseline, df = 1.02)).

this transition, the discriminatory ability of the model declines over time until 36 months. Finally, for the transition Entrance → Transfer to haemodialysis, estimates of time-varying AUC approach 0.50 and increase slightly after approximately 36 months.

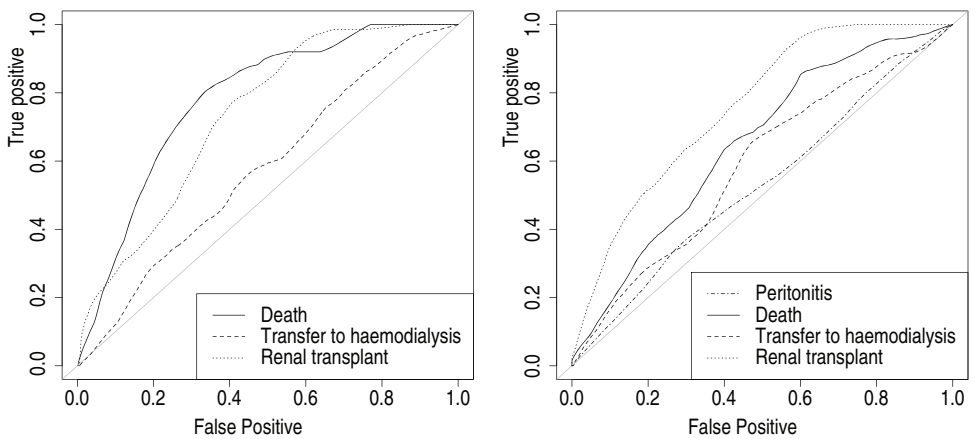


Figure 4 Estimated ROC curves ROC(t) at $t = 36$ months; (a) for transitions Entrance → Death, Entrance → Transfer to haemodialysis and Entrance → Renal transplant; (b) for transitions Entrance → Peritonitis, Peritonitis → Death, Peritonitis → Transfer to haemodialysis and Peritonitis → Renal transplant.

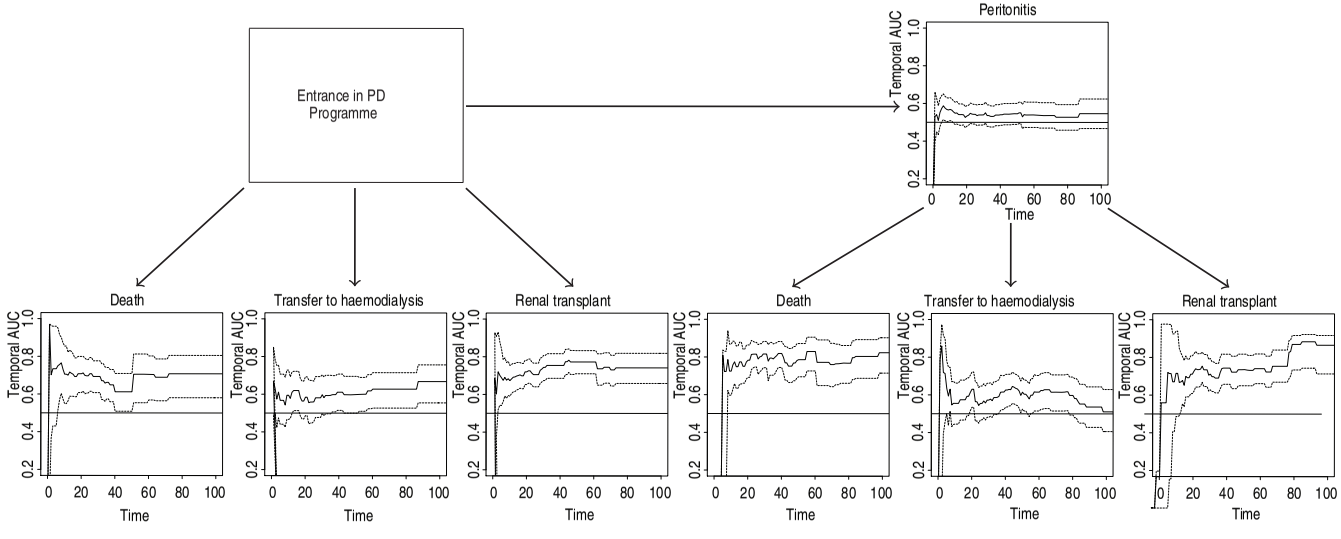


Figure 5 Temporal AUC for STAR model (solid line) and bootstrap 95% confidence intervals (dotted line).

5 Discussion

In this article, we analyze the use of time-dependent ROC methodology to evaluate the predictive accuracy of semiparametric multi-state models in the presence of competing risks. This analysis was performed to evaluate the trajectory of PD patients between the entrance in the PD programme until the occurrence of one of the following events: death, transfer to haemodialysis or renal transplantation. Since we have more than one absorbing event, we are in the presence of a competing risks problem.

When in the PD programme, a patient can experience one or more peritonitis episodes. Though we chose this transient state due to its clinical relevance and its high frequent occurrence, other relevant transient states could have also been analyzed, pointing to the relevance of the multi-state analysis approach in this context. For example, hospitalization, anuria or other major complications, that condition patient outcomes, could have also been considered. Given the complexity of these data structures, we considered the final model representation as a nested series of competing risks experiments (Beyersmann *et al.*, 2012), thus benefiting from the advantages of these models, including the Markov assumption.

In this study, we performed a Cox-type structured hazard multi-state regression model in order to identify factors associated with the different hazards transitions. In this methodology, the transition intensities were specified in a multiplicative manner allowing the inclusion of flexible non-parametric effects and also time-varying effects of categorical covariates, which are both main advantages of this method. Flexible HR curves allowed a simple and intuitive interpretation of results obtained with STAR models, particularly time-varying and non-parametric effects of some categorical covariates along transitions. We studied one model including the parametric effects of gender, diabetes and previous renal replacement therapy and the non-parametric effects of age and the (log) baseline hazard rate. Both diabetes and previous renal replacement therapy (non-naive PD status) impact significantly on the hazard of death but not on the hazard of transition to haemodialysis. When a first peritonitis occurs, results remain similar. All covariates considered in this study were available for the total sample. More covariates could not be used because of many missing values. For this reason, it is necessary to alert clinicians for the implications and advantages of a proper data collection to perform a correct and detailed data analysis.

The results produce new information about the PD programme. For instance, the clear evidence that diabetes has a significant time-varying effect being associated with a significantly higher hazard of transition to death state with higher time under therapy, and that age is not associated with higher hazard of transition to haemodialysis whether peritonitis occurs or not. These are clinically relevant inputs that stress the relevance of the statistical methodologies here discussed. It should be noticed, however, that other data structures and covariates must be evaluated in order to obtain more accurate knowledge about the trajectory of PD patients, since a first peritonitis event might not be as clinically relevant as the cumulative number of peritonitis, peritonitis severity or other metabolic complications more hardly quantifiable.

Time-dependent ROC curves (and corresponding AUCs) proved to be useful tools in analyzing the predictive accuracy of multi-state analysis in the presence of competing risks. A key advantage of the ROC curve is that changing the units in which a marker is measured has no impact on it. So, in the multi-state framework, the ROC curve provides a natural common scale for comparing different risk scores among different transitions, even when they are measured in completely different units.

Thus, the analysis of time-varying AUC allowed us to conclude that the considered STAR model presents a high degree of validity for some transitions, namely: Entrance $\rightarrow$ Death, Entrance $\rightarrow$ Renal transplant, Peritonitis $\rightarrow$ Death and Peritonitis $\rightarrow$ Renal transplant. The decline of the discriminatory ability of the model over time for the transition Entrance $\rightarrow$ Death may be related to the increase of comorbidities associated with disease and therapy, which were not considered in this model. Then, these facts suggest that other predictive factors, such as cardiovascular comorbidities, loss of residual renal function, volemic control or systematic inflammation, which may be evaluated by continuous clinical parameters, need to be considered in future studies, specifically when the analysis of technique survival is concerned (i.e., when the final endpoint is transfer to haemodialysis).

In this article, we adopted the definition of the time-dependent ROC proposed by Heagerty *et al.* (2000). This definition (see Section 3) refers to the evaluation of the prediction accuracy of a model score (based on covariates measured at baseline) to distinguish between subjects having an event before time t, from those who do not. As already commented in the Introduction, an alternative definition might have been used as well. Such an alternative scenario arises when clinical interest focuses on the correct classification of subjects at time t among those who are still at risk (Heagerty and Zheng, 2005). Though it is outside the scope of this article, it could be worth extending such definition to the multi-state framework.

Also, it should be noted that other approaches using ROC curves have been proposed very recently to evaluate the predictive accuracy of regression models in the presence of competing risks (Blanche *et al.*, 2013; Saha and Heagerty, 2010; Zheng *et al.*, 2012). However, these approaches appear to be more appropriate in the classical survival analysis in the presence of competing risks, for example when Cox cause-specific hazard regression model (Prentice *et al.*, 1978) and Fine and Gray regression model (Fine and Gray, 1999) are considered.

It is noteworthy that in medical practice, more general specifications of the predictor are needed in the structured Cox-type hazard multi-state regression framework. Such specifications are related to the inclusion of interactions between continuous covariates, and by extension of time-varying effects of continuous covariates, which allow relaxing the proportional hazards assumption at each transition. A possible solution to this problem would be based on the equivalence in terms of likelihood between piecewise exponential models obtained via data augmentation and STAR models or a flexible alternative based on P-spline such as generalized linear additive smooth structures (GLASS) model (Eilers and Marx, 2002) with Poisson error structure (Fahrmeir and Kneib, 2011; Rodriguez-Girondo *et al.*, 2013). In this case, the time axis is partitioned and the hazard is assumed to be constant over each interval. In this way, the estimation algorithm allows for all types

of effects, including combinations of non-linear and time-varying effects, and hence, HR surfaces can be derived. We are currently researching such possible extensions. In summary, our model is an informative tool for the medical decision process and evaluation of the patients in PD. The use of STAR models complemented with the use of time-dependent ROC curves in this context allows the identification of relevant factors associated with specific transitions. The identification of these factors, which could not have been obtained with standard survival models, contributes to a better knowledge of patients' trajectories, and consequently, in better management of the treatment programme.

ACKNOWLEDGEMENTS

The authors would like to thank Thomas Kneib for the helpful suggestions. This work was partially supported by grants MTM2011-28285-C02-01 and MTM2011-15849-E from the Spanish Ministry of Science and Innovation.

References

Abraham G, Kumar V, Nayak KS, Ravichandran R, Srinivasan G, Krishnamurthy M, Prasath AK, Kumar S, Thiagarajan T, Mathew M and Lesley N (2010) Predictors of long-term survival on peritoneal dialysis in south india: A multicenter study. *Peritoneal Dialysis International*, **30**, 29–34. doi: 10.3747/pdi.2008.00028

Akritas MG (1994) Nearest-neighbor estimation of a bivariate distribution under random censoring. *Annals of Statistics*, **22**, 1299–327. doi: DOI 10.1214/aos/1176325630

Andersen PK and Keiding N (2012) Interpretability and importance of functionals in competing risks and multistate models. *Statistics in Medicine*, **31**, 1074–88. doi: 10.1002/sim.4385

Belitz C, Brezger A, Kneib T, Lang S and Umlauf N (2012) Bayesx - software for bayesian inference in structured additive regression models. Version 3.0. URL: http://www.bayesx.org

Beyersmann J, Schumacher M and Allignol A (2012) *Competing risks and multistate models with R*. Use R! New York: Springer.

Blanche P, Dartigues JF and Jacqmin-Gadda H (2013) Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in Medicine*, **32**, 5381–97.

Brezger A and Lang S (2006) Generalized structured additive regression based on bayesian P-splines. *Computational Statistics & Data Analysis*, **50**, 967–91. doi: DOI 10.1016/j.csda.2004.10.011

Cadarso-Suárez C, Meira-Machado L, Kneib T and Gude F (2010) Flexible hazard ratio curves for continuous predictors in multistate models: An application to breast cancer data. *Statistical Modelling*, **10**, 291–314. doi: 10.1177/1471082x0801000303

Chen HC, Kodell R, Cheng K and Chen J (2012) Assessment of performance of survival prediction models for cancer prognosis. *BMC Medical Research Methodology*, **12**, 102.

Davenport A (2009) Peritonitis remains the major clinical complication of peritoneal dialysis: The London, UK, peritonitis audit 2002–2003. *Peritoneal Dialysis International*, **29**, 297–302.

Davison A and Hinkley D (1997) *Bootstrap Methods and Their Application*. UK: Cambridge University Press. ISBN 9780521574716

de Wreede LC, Fiocco M and Putter H (2011) Mstate: An r package for the analysis of competing risks and multi-state models. *Journal of Statistical Software*, **38**, 1–30.

Efron B (1981) Censored data and the bootstrap. *Journal of the American Statistical Association*, **76**, 312–19. doi: 10.1080/01621459.1981.10477650

Eilers PHC and Marx BD (1996) Flexible smoothing with b-splines and penalties. *Statistical Science*, **11**, 89–102. doi: DOI 10.1214/ss/1038425655

Eilers PHC and Marx BD (2002) Generalized linear additive smooth structures. *Journal of Computational and Graphical Statistics*, **11**, 758–83.

Etzioni R, Pepe M, Longton G, Hu C and Goodman G (1999) Incorporating the time dimension in receiver operating characteristic curves: A case study of prostate cancer. *Medical Decision Making*, **19**, 242–51.

Fahrmeir L and Kneib T (2011) *Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data*. Oxford, UK: Oxford University Press. ISBN 9780199533022.

Fine JP and Gray RJ (1999) A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, **94**, 496–509.

Gooley TA, Leisenring W, Crowley J and Storer BE (1999) Estimation of failure probabilities in the presence of competing risks: New representations of old estimators. *Statistics in Medicine*, **18**, 695–706.

Hastie T and Tibshirani R (1990a) Exploring the nature of covariate effects in the proportional hazards model. *Biometrics*, **46**, 1005–16.

—— (1990b) *Generalized additive models* (Vol. 43). Boca Raton, FL: CRC Press.

Heagerty PJ, Lumley T and Pepe MS (2000) Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*, **56**, 337–44.

Heagerty PJ and Zheng Y (2005) Survival model predictive accuracy and ROC curves. *Biometrics*, **61**, 92–105.

Hennerfeind A, Brezger A and Fahrmeir L (2006) Geoadditive survival models. *Journal of the American Statistical Association*, **101**, 1065–75.

Hougaard P (1999) Multi-state models: A review. *Lifetime Data Analysis*, **5**, 239–64

Kneib T and Hennerfeind A (2008) Bayesian semiparametric multi-state models. *Statistical Modelling*, **8**, 169–98.

Kneib T and Fahrmeir L (2007) A mixed model approach for geoadditive hazard regression. *Scandinavian Journal of Statistics*, **34**, 207–28.

Kotsanas D, Polkinghorne KR, Korman TM, Atkins RC and Brown F (2007) Risk factors for peritoneal dialysis-related peritonitis: Can we reduce the incidence and improve patient selection? *Nephrology*, **12**, 239–45.

Levey AS, Andreoli SP, DuBose T, Provenzano R and Collins AJ (2007) Chronic kidney disease: Common, harmful, and treatable—World Kidney Day 2007. *Journal of the American Society of Nephrology*, **18**, 374–78.

Martins M, Rodrigues A, Pedrosa JM, Carvalho MJ, Cabrita A and Oliveira R (2013) Update on the challenging role of biofilms in peritoneal dialysis. *Biofouling*, **29**, 1015–27.

Martinussen T and Scheike T (2010) *Dynamic regression models for survival data*. USA: Springer. ISBN 9781441919045.

Meira-Machado L, de Una-Alvarez J, Cadarso-Suárez C and Andersen PK (2009) Multi-state models for the analysis of time-to-event data. *Statistical Methods in Medical Research*, **18**, 195–222.

Meira-Machado L, Cadarso-Suárez C, Gude F and Arajo A (2013) Smooth HR: An R package for pointwise nonparametric estimation of hazard ratio curves of continuous predictors. *Computational and Mathematical Methods in Medicine*, **28**, 1–11.

Pepe M (2004) *The statistical evaluation of medical tests for classification and prediction*. USA: Oxford University Press. ISBN 9780198565826.

Pepe M, Zheng Y, Jin Y, Huang Y, Parikh C and Levy W (2008) Evaluating the ROC

performance of markers for future events. *Lifetime Data Analysis*, **14**, 86–114.

Perl J, Wald R, Bargman JM, Na Y, Jassal SV, Jain AK, Moist L and Nessim SJ (2012) Changes in patient and technique survival over time among incident peritoneal dialysis patients in Canada. *Clinical Journal of the American Society of Nephrology*, **7**, 1145–54.

Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT and Breslow NE (1978) Analysis of failure times in presence of competing risks. *Biometrics*, **34**, 541–54.

Putter H, Fiocco M and Geskus RB (2007) Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*, **26**, 2389–430.

R Development CoreTeam (2013) R: A language and environment for statistical computing. Vienna, Austria: The R Foundation for Statistical Computing. ISBN: 3-900051-07-0. URL: http://www.R-project.org/

Rocha A, Rodrigues A, Teixeira L, Carvalho MJ, Mendonça D and Cabrita A (2012) Temporal trends in peritonitis rates, microbiology and outcomes: The major clinical complication of peritoneal dialysis. *Blood Purification*, **33**, 284–91.

Rodriguez-Girondo M, Kneib T, Cadarso-Suárez C and Abu-Assi E (2013) Model building in nonproportional hazard regression. *Statistics in Medicine*, **32**, 5301–14.

Saha P and Heagerty PJ (2010) Time-dependent predictive accuracy in the presence of competing risks. *Biometrics*, **66**, 999–1011.

Schemper M (2003) Predictive accuracy and explained variation. *Statistics in Medicine*, **22**, 2299–2308.

Swets J and Pickett R (1982) *Evaluation of diagnostic systems: methods from signal detection theory*. USA: Elsevier Science & Technology Books. ISBN 9780126790801.

Teixeira L, Rodrigues A, Carvalho MJ, Cabrita A and Mendonça D (2013) Modelling competing risks in nephrology research: An example in peritoneal dialysis. *BMC Nephrology*, **14**, 110.

Yang X, Yi C, Liu X, Guo Q, Yang R, Cao P, Lin J, Mao H and Yu X (2013) Clinical outcome and risk factors for mortality in chinese patients with diabetes on peritoneal dialysis: A 5-year clinical cohort study. *Diabetes Research and Clinical Practice*, **100**, 354–61.

Zheng YY, Cai TX, Jin YY and Feng ZD (2012) Evaluating prognostic accuracy of biomarkers under competing risk. *Biometrics*, **68**, 388–96.

Zhou X, Obuchowski N and McClish D (2011) *Statistical Methods in Diagnostic Medicine*. USA: Wiley. ISBN 9780470906507.