SORT 40 (2) July-December 2016, 375-398

Smoothed landmark estimators of the transition probabilities

Luís Meira-Machado

Abstract

One important goal in clinical applications of multi-state models is the estimation of transition probabilities. Recently, landmark estimators were proposed to estimate these quantities, and their superiority with respect to the competing estimators has been proved in situations in which the Markov condition is violated. As a weakness, it provides large standard errors in estimation in some circumstances. In this article, we propose two approaches that can be used to reduce the variability of the proposed estimator. Simulations show that the proposed estimators may be much more efficient than the unsmoothed estimator. A real data illustration is included.

MSC: 62N02, 62G05, 62P10.

Keywords: Kaplan-Meier, Multi-state model, Nonparametric estimation, Presmoothing, Survival Analysis.

1. Introduction

The analysis of survival data may be described by the Markov process considering the transition from an initial 'alive' state to a single ultimate state or endpoint 'dead'. However, in most longitudinal medical studies more than one endpoint can be defined. In breast cancer trials, for instance, several endpoints, such as disease-free survival, local recurrence, distant metastasis or death are possible. Multi-state models are a useful way of describing such a process in which an individual moves through a number of finite states in continuous time. A wide range of medical situations have been modeled using multi-state methods, for example, HIV infection and AIDS (Gentleman et al., 1994), liver cirrhosis (Andersen et al., 2002), breast cancer (Pérez-Ocón et al., 2001; Putter et al., 2007) and problems following heart transplantation (Meira-Machado et al., 2009). A commonly-used model is the illness-death model, with three states representing health, illness and death (Figure 1). Individuals start in the healthy state and subsequently move

Centre of Mathematics and Department of Mathematics and Applications, University of Minho, Guimarães, Portugal. lmachado@math.uminho.pt

Received: March 2016
Accepted: November 2016

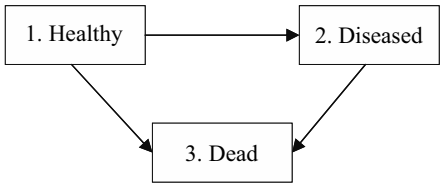

Figure 1: *Illness-death model.*

either to the diseased state or to the dead absorbing state. In the irreversible version of this model, individuals in the diseased state will eventually move to the dead state without any possibility of recovery. Methods developed for the progressive illness-death model have a wide range of applications in longitudinal medical studies.

One important feature of multi-state models is their ability to obtain predictions of the clinical prognosis of a patient at a certain point in his/her recovery or illness process. Various aspects of the model dynamics can be captured through the transition probabilities. Traditionally, the transition probabilities are estimated using Aalen-Johansen estimator (Aalen and Johansen, 1978) which assumes the process to be Markovian. Substitute estimators for the Aalen-Johansen estimator for a general non-Markov illness-death process without recovery were introduced by Meira-Machado et al. (2006). These authors showed that the new estimators may behave much more efficiently than the Aalen-Johansen when the Markov assumption does not hold. This work has been revisited by Allignol et al. (2014), who proposed a closely related non-Markov estimator too. However, both of the proposed non-Markov estimators have the drawback of requiring that the support of the censoring distribution contains the support of the lifetime distribution, which is not often the case. To avoid this problem, de Uña-Álvarez and Meira-Machado (2015) propose new estimation methods which are consistent regardless the Markov condition and the referred assumption about the censoring support. The idea behind the proposed methods is to use specific subsamples or portions of data at hand (namely, those observed to be in a given state at a pre-specified time point). Such an approach is known in the literature as the landmark methodology (van Houwelingen, 2007). Simulations reported in the paper by de Uña-Álvarez and Meira-Machado (2015) reveal significant improvements on the behaviour of the new method. For small sample sizes and/or large proportion of censored data the landmark approach may result in a wiggly estimator with fewer jump points. This will be more prominent is some transition probabilities. To avoid this problem, we propose two approaches that can be used to reduce the variability of the landmark estimator. A simple approach is based on spline smoothing. Another valid approach is to consider a modification of the landmark estimator based on presmoothing (Dikta, 1998). Simulation studies reported in Section 3 show that the proposed estimators may be much more efficient than the completely nonparametric estimator. In addition, we introduce nonparametric estimators based on the landmark approach that account for the influence of covariates in the transition probabilities.

The organization of the paper is as follows. In Section 2, we introduce the notation and revisit the estimator proposed by de Uña-Álvarez and Meira-Machado (2015). New smoothed estimators are also introduced. The performance of the three sets of estimators is investigated through simulations in Section 3, while in Section 4 the methods are compared through the analysis of medical data from a clinical trial on breast cancer from Germany. In Section 5 we give a brief overview of the R package developed by the authors. Main conclusions are reported in Section 6.

2. Transition probabilities

2.1. Notation and preliminaries

A multi-state model is a stochastic process $(Y(t), t \in \mathscr{T})$ with a finite state space in continuous time. These models are a useful way of describing a process in which an individual moves through a series of states. In this paper, we consider the progressive illness-death model depicted in Figure 1 and we assume that all the subjects are in State 1 at time $t = 0$. Extensions to progressive processes beyond the three-state illness-death model can also be considered following the ideas given in the paper by de Uña-Álvarez and Meira-Machado (2015) (Section 5).

The progressive illness-death model is characterized by the three random variables T_{ij}, $1 \leq i < j \leq 3$, that represent the potential transition times from State i to State j. According to this notation, subjects not visiting State 2 will reach State 3 at time T_{13}. This time will be $T_{12} + T_{23}$ if he/she passes through State 2 before, where the variables T_{12} and T_{23} are recorded successively, rather than simultaneously. In this model we have two competing transitions leaving State 1. Therefore, we denote by $\rho = I(T_{12} \leq T_{13})$ the indicator of visiting State 2 at some time, $Z = \min(T_{12}, T_{13})$ the sojourn time in State 1, and $T = Z + \rho T_{23}$ the total survival time of the process. This means that $\rho = I(Z < T)$.

As usual, assume that these event times are subject to univariate right-censoring denoted by C, which we assume to be independent of (Z,T). Define $\widetilde{Z} = \min(Z,C)$ and $\widetilde{T} = \min(T,C)$ for the censored versions of Z and T. Then, put $\Delta_1 = I(Z \leq C)$ and $\Delta = I(T \leq C)$ for the respective censoring indicators. Finally, the available data is $(\widetilde{Z}_i, \widetilde{T}_i, \Delta_{1i}, \Delta_i)$, $1 \leq i \leq n$, iid copies of $(\widetilde{Z}, \widetilde{T}, \Delta_1, \Delta)$.

In the illness-death model, the target is each of the five transition probabilities $p_{ij}(s,t) = P(Y(t) = j | Y(s) = i)$, where $1 \leq i \leq j \leq 3$ and $s \leq t$ are two pre-specified time points. However, since we have two obvious relations, $p_{12}(s,t) = 1 - p_{11}(s,t) - p_{13}(s,t)$ and $p_{22}(s,t) = 1 - p_{23}(s,t)$, in practice one only need to estimate three of these quantities. According to our notations, the transition probabilities are written as

$$p_{11}(s,t) = P(Z > t \mid Z > s), \quad p_{12}(s,t) = P(Z \leq t, T > t \mid Z > s),$$

$$p_{13}(s,t) = P(T \leq t \mid Z > s), \quad p_{22}(s,t) = P(Z \leq s, T > t \mid Z \leq s, T > s),$$
$$p_{23}(s,t) = P(T \leq t \mid Z \leq s, T > s). \tag{1}$$

2.2. Landmark estimators

According to the landmark approach (van Houwelingen, 2007) nonparametric estimators for the transition probabilities can be introduced by considering specific subsamples or portions of the data. For example, given the time point s, to estimate $p_{1j}(s,t)$ for $j = 1,2,3$ the analysis can be restricted to the individuals observed in State 1 at time s. This set is just $\mathscr{S}_1 = \left\{ i : \widetilde{Z}_i > s \right\}$. As explained in de Uña-Álvarez and Meira-Machado (2015) as long as C is independent of Z, a subject in $\mathscr{S}_1$ is representative of those individuals for which Z exceeds s. On the other hand, for the subpopulation $\widetilde{Z} > s$, the censoring time C is still independent of the pair (Z,T) and, therefore, Kaplan-Meier-based estimation will be consistent. The same applies to the analysis restricted to the individuals observed in State 2 at time s, say $\mathscr{S}_2 = \left\{ i : \widetilde{Z}_i \leq s < \widetilde{T}_i \right\}$, which serves to introduce landmark estimators for $p_{2j}(s,t)$, $j = 2,3$.

The transition probability $p_{11}(s,t)$ is defined as the survival function at time t, among the individuals observed in State 1 at time s, which can be estimated by the ordinary Kaplan-Meier estimator (Kaplan and Meier, 1958) of the sojourn time distribution in State 1, based on the pairs $(\widetilde{Z}_i, \Delta_{1i})$'s in the subsample $\mathscr{S}_1$. Similarly, the transition probability $p_{13}(s,t)$ is defined as one minus the survival function (of the total time) at time t in the same subset $\mathscr{S}_1$. The transition probability $p_{23}(s,t)$ is defined as one minus the survival function (of the total time) at time t in the subset $\mathscr{S}_2$. The landmark estimators given in the paper by de Uña-Álvarez and Meira-Machado (2015) are defined in terms of multivariate 'Kaplan-Meier integrals' with respect to the marginal distribution of the first time, for the transition probability $p_{11}(s,t)$, and with respect to the marginal distribution of the total time T in the remaining transitions.

To formally present the estimators, we need to introduce the expressions for the Kaplan-Meier weights: $w_i^{(s)}$ - the Kaplan-Meier weights attached to $\widetilde{Z}_i$ when estimating the marginal distribution of Z from the $(\widetilde{Z}_i, \Delta_{1i})$'s in subset $\mathscr{S}_1$, $W_i^{(s)}$ - the Kaplan-Meier weights attached to $\widetilde{T}_i$ when estimating the marginal distribution of T from the $(\widetilde{T}_i, \Delta_i)$'s in subset $\mathscr{S}_1$, and $W_i^{[s]}$ - the Kaplan-Meier weights attached to $\widetilde{T}_i$ when estimating the marginal distribution of T from the $(\widetilde{T}_i, \Delta_i)$'s in subset $\mathscr{S}_2$. Let $\left(\widetilde{Z}_{(i)}^{(s)}, \Delta_{1[i]}^{(s)}\right)$, $i = 1,...,n_{1s}$, be the $\left(\widetilde{Z}, \Delta_1\right)$-sample in $\mathscr{S}_1$ ordered with respect to $\widetilde{Z}$, and $\left(\widetilde{T}_{(i)}^{(s)}, \Delta_{[i]}^{(s)}\right)$, $i = 1,...,n_{1s}$, be the $\left(\widetilde{T}, \Delta\right)$-sample in $\mathscr{S}_1$ ordered with respect to $\widetilde{T}$. Then,

$$\widehat{p}_{11}^{ldm}(s,t) = 1 - \sum_{i=1}^{n_{1s}} w_i^{(s)} I(\widetilde{Z}_{(i)}^{(s)} \leq t), \tag{2}$$

$$\widehat{p}_{13}^{ldm}(s,t) = \sum_{i=1}^{n_{1s}} W_i^{(s)} I(\widetilde{T}_{(i)}^{(s)} \leq t), \tag{3}$$

where

$$w_i^{(s)} = \frac{\Delta_{1[i]}^{(s)}}{n_{1s}-i+1} \prod_{j=1}^{i-1} \left[1 - \frac{\Delta_{1[j]}^{(s)}}{n_{1s}-j+1} \right], \quad 1 \leq i \leq n_{1s};$$

and

$$W_i^{(s)} = \frac{\Delta_{[i]}^{(s)}}{n_{1s}-i+1} \prod_{j=1}^{i-1} \left[1 - \frac{\Delta_{[j]}^{(s)}}{n_{1s}-j+1} \right], \quad 1 \leq i \leq n_{1s}.$$

Similarly, one can introduce the corresponding estimator for $p_{23}(s,t)$. Let $\left(\widetilde{T}_{(i)}^{[s]}, \Delta_{[i]}^{[s]}\right)$, $i = 1, ..., n_{2s}$, is the $\left(\widetilde{T}, \Delta\right)$-sample in $\mathscr{S}_2$ ordered with respect to $\widetilde{T}$. Then,

$$\widehat{p}_{23}^{ldm}(s,t) = \sum_{i=1}^{n_{2s}} W_i^{[s]} I(\widetilde{T}_{(i)}^{[s]} \leq t), \tag{4}$$

where

$$W_i^{[s]} = \frac{\Delta_{[i]}^{[s]}}{n_{2s}-i+1} \prod_{j=1}^{i-1} \left[1 - \frac{\Delta_{[j]}^{[s]}}{n_{2s}-j+1} \right], \quad 1 \leq i \leq n_{2s}.$$

The estimators $\widehat{p}_{ij}^{ldm}(s,t)$ have the simple form of a Kaplan-Meier estimator, based on a certain subsample which is determined by the time point s. Thus, they can also be expressed in the form of inverse of probability censoring weighted average (IPCW) (Satten and Datta, 2001),

$$\widehat{p}_{11}^{ldm}(s,t) = 1 - \frac{1}{n_{1s}} \sum_{i=1}^{n_{1s}} \frac{I(\widetilde{Z}_{(i)}^{(s)} \leq t)\Delta_{1[i]}^{(s)}}{\widehat{G}(\widetilde{Z}_{(i)}^{(s)})},$$

$$\widehat{p}_{13}^{ldm}(s,t) = \frac{1}{n_{1s}} \sum_{i=1}^{n_{1s}} \frac{I(\widetilde{T}_{(i)}^{(s)} \leq t)\Delta_{[i]}^{(s)}}{\widehat{K}_1(\widetilde{T}_{(i)}^{(s)})},$$

and

$$\widehat{p}_{23}^{ldm}(s,t) = \frac{1}{n_{2s}} \sum_{i=1}^{n_{2s}} \frac{I(\widetilde{T}_{(i)}^{[s]} \leq t)\Delta_{[i]}^{[s]}}{\widehat{K}_2(\widetilde{T}_{(i)}^{[s]})},$$

where G, K_1 and K_2 are the survival functions of the censoring variable C, which can be consistently estimated by the Kaplan-Meier approach considering events as 'censored' observations and censored observations as 'events'. Here, $\widehat{G}$ stands for the Kaplan-Meier estimator (of the censoring survival function) based on the $(\widetilde{Z}_i, 1-\Delta_{1i})$'s in subset $\mathscr{S}_1$; whereas, $\widehat{K}_1$ and $\widehat{K}_2$ stand for the Kaplan-Meier estimator (of the censoring survival function) based on the $(\widetilde{T}_i, 1-\Delta_i)$'s in subset $\mathscr{S}_1$ and $\mathscr{S}_2$, respectively.

It is important to mention that $\widehat{p}_{11}^{ldm}(s,t)$ is equivalent to the estimator given by Meira-Machado et al. (2006) and the so-called Aalen-Johansen estimator (Aalen and Johansen, 1978) of $p_{11}(s,t)$, which is consistent regardless of the Markov assumption. In addition, for $s=0$, the landmark estimators are known as the occupation probabilities and they are equivalent to those provided by Meira-Machado et al. (2006).

2.3. Smooth landmark estimators

The standard error of the landmark estimators introduced in the previous subsection may be large when the censoring is heavy, particularly with a small sample size. This problem may be more obvious when estimating the transition probabilities $p_{ij}(s,t)$ for large values of s. In this section, we propose two smoothed versions of the nonparametric landmark estimators given in the previous subsection. One simple approach is based on the use of constrained penalized regression splines (Meyer, 2008, 2012; Wood, 2006). We also introduce a semiparametric estimator which uses a presmoothed version of the Kaplan-Meier estimator (Dikta, 1998; Jácome and Iglesias, 2008; López-de-Ullibarri and Jácome, 2013) pertaining to the distribution of the survival times to weight the data.

2.3.1. Constrained penalized splines

Constrained penalized regression splines can be used as a simple approach which provides smooth estimation of the transition probabilities. These methods can be used under some constraints of shape, such as monotonicity (required for the transition probabilities $p_{11}(s,t)$, $p_{13}(s,t)$, $p_{22}(s,t)$ and $p_{23}(s,t)$) and to force a fit curve to go through a particular point. The later constraint is also important since for $s=t$ obvious conditions are required ($p_{11}(s,s)=p_{22}(s,s)=1$ and $p_{12}(s,s)=p_{13}(s,s)=p_{23}(s,s)=0$). To obtain spline-based landmark estimators, $\widehat{p}_{ij}^{crs}(s,t)$, we propose the use of the cubic regression splines.

The key assumption underlying regression spline smoothing is that, for a fixed value of s, the unknown functions $p_{ij}(s,t)$ can be approximated by polynomial splines, defined

on a set of knots (join points) within the domain of $A = [s, \tau_T]$ where τ_T is the upper bound of the support of T. For a fixed value of s, we first obtain the (landmark) estimates of the transition probabilities over all possible time values t with $s < t$, and then, define a cubic spline basis defined by a modest sized set of knots spread evenly through the interval $A = [s, \tau_T]$. For each transition probability $p_{ij}(s,t)$ the use of regression splines provide one approach that allows flexible relationships between a covariate X (time values in A) and the average response (i.e., the landmark estimates of the transition probabilities) as a function of the variable X.

Cubic spline functions are piecewise continuous curves defined by polynomial functions of degree 3. These functions are built joining the piecewise functions on equally spaced join points (also known as knots) so that they are continuous in value, as well as its first two derivatives. This is done by choosing a cubic regression spline basis for which many alternatives can be found (see for example Durrleman and Simon (1989) or Wood (2006)). One approach is to parameterize the spline terms of its values at the knots (Wood, 2006). Specifically, we can write the spline function as a function of $u \in A$

$$f(u) = \sum_{i=1}^{q} \delta_i b_i(u)$$

where q is the number of knots, the b_i are the basis functions of at least second order for representing smooth functions over a given interval, and the δ_i are the spline coefficients. Constraints forcing the curve to pass through a specific point can be imposed. This can be done by creating a regression spline basis, making sure there is a knot at the constraint point. Monotonicity constraints can also be imposed. Penalization is achieved by the conventional integrated square second derivative cubic spline penalty. In practice, we use the gam function in the R package mgcv (Wood, 2006) to obtain the transition probability curves as well for obtaining the predicted values of the smooth curves at the new values. The mgcv implementation of gam, by default uses basis functions for these splines that are designed to be optimal, given the number basis functions used. For details about these methods, see for example Wood (2006) or Pya and Wood (2015).

2.3.2. Presmoothed estimators

The variance of the landmark estimators may also be reduced by presmoothing. Successful applications of presmoothed estimators include estimation of the survival function (Dikta, 1998; Meira-Machado et al., 2016), nonparametric curve estimation (Cao and Jácome, 2004), regression analysis (de Uña-Álvarez and Rodríguez-Campos, 2004; Jácome and Iglesias, 2010), estimation of the bivariate distribution of censored gap times (de Uña-Álvarez and Amorim, 2011), and the estimation of the transition probabilities (Amorim et al., 2011; Moreira et al., 2013). All these references concluded that the presmoothed estimators have improved variance when compared to purely nonparamet-

ric estimators. In this paper, we show that presmoothing is also useful to improve efficiency of the landmark estimators introduced in a previous section. This 'presmoothing' is obtained by replacing the censoring indicator variables in the expression of the Kaplan-Meier weights by a smooth fit. This preliminary smoothing may be based on a certain parametric family such as the logistic, or on a nonparametric estimator of the binary regression curve. When the parametric family is the right one, parametric presmoothing (Dikta, 1998) leads to more efficient estimation than that associated to the unsmoothed estimator. Nonparametric presmoothing (Cao et al., 2005) is useful when there is a clear risk of a miss-specification of the parametric model. The validity of a given parametric model for presmoothing can be checked graphically or formally, by applying a goodness-of-fit test. In this paper we consider estimators obtained using standard logistic regression. The corresponding (semiparametric) presmoothed landmark estimators of the transition probabilities are given by

$$\widehat{p}_{11}^{prs}(s,t) = 1 - \sum_{i=1}^{n_{1s}} pw_i^{(s)} I(\widetilde{Z}_{(i)}^{(s)} \leq t), \tag{5}$$

$$\widehat{p}_{13}^{prs}(s,t) = \sum_{i=1}^{n_{1s}} PW_i^{(s)} I(\widetilde{T}_{(i)}^{(s)} \leq t), \tag{6}$$

and

$$\widehat{p}_{23}^{prs}(s,t) = \sum_{i=1}^{n_{2s}} PW_i^{[s]} I(\widetilde{T}_{(i)}^{[s]} \leq t) \tag{7}$$

where the presmoothed Kaplan-Meier weights are defined as follow:

$$pw_i^{(s)} = \frac{m_{0n}(\widetilde{Z}_{(i)}^{(s)})}{n_{1s}-i+1} \prod_{j=1}^{i-1} \left[1 - \frac{m_{0n}(\widetilde{Z}_{(i)}^{(s)})}{n_{1s}-j+1} \right], \quad 1 \leq i \leq n_{1s},$$

$$PW_i^{(s)} = \frac{m_n(\widetilde{Z}_{[i]}^{(s)}, \widetilde{T}_{(i)}^{(s)})}{n_{1s}-i+1} \prod_{j=1}^{i-1} \left[1 - \frac{m_n(\widetilde{Z}_{[i]}^{(s)}, \widetilde{T}_{(i)}^{(s)})}{n_{1s}-j+1} \right], \quad 1 \leq i \leq n_{1s},$$

and

$$PW_i^{[s]} = \frac{m_n(\widetilde{Z}_{[i]}^{[s]}, \widetilde{T}_{(i)}^{[s]})}{n_{2s}-i+1} \prod_{j=1}^{i-1} \left[1 - \frac{m_n(\widetilde{Z}_{[i]}^{[s]}, \widetilde{T}_{(i)}^{[s]})}{n_{2s}-j+1} \right], \quad 1 \leq i \leq n_{2s},$$

where $\left(\widetilde{Z}_{[i]}^{(s)}, \widetilde{T}_{(i)}^{(s)}, \Delta_{[i]}^{(s)}\right)$, $i = 1, ..., n_{1s}$, is the $\left(\widetilde{Z}, \widetilde{T}, \Delta\right)$-sample in $\mathscr{S}_1$ ordered with respect to $\widetilde{T}$, and $\left(\widetilde{Z}_{[i]}^{[s]}, \widetilde{T}_{(i)}^{[s]}, \Delta_{[i]}^{[s]}\right)$, $i = 1, ..., n_{2s}$, is the $\left(\widetilde{Z}, \widetilde{T}, \Delta\right)$-sample in $\mathscr{S}_2$ ordered with respect to $\widetilde{T}$. Here, $m_{0n}(u)$ and $m_n(u,v)$ stand for estimators of the binary regression functions $m_0(u) = P(\Delta_1^{(s)} = 1 \mid \widetilde{Z}^{(s)} = u)$ and $m(u,v) = P(\Delta^{(s)} = 1 \mid \widetilde{Z}^{(s)} = u, \widetilde{T}^{(s)} = v)$, respectively. In this work we assume that these functions belong to a parametric (smooth) family of binary logistic regression curves. For example for $m_{0n}(u)$, we assume that $m_{0n}(u) = m(u;\beta)$ where β is a vector of parameters which typically will be computed by maximizing the conditional likelihood of the $\Delta_1^{(s)}$'s given $\widetilde{Z}^{(s)}$.

As discussed in Amorim et al. (2011) the function $m(u,v)$ will typically be discontinuous along the line $v = u$, that is, for those covariate values $(\widetilde{Z}, \widetilde{T})$ corresponding to individuals who are censored while being in state 1 or who suffer a direct transition to the absorbing state. In order to construct $m_n(u,v)$ we use the ideas proposed by Amorim et al. (2011).

Note that, unlike the unsmoothed landmark estimators, the presmoothed estimators can attach positive mass to pair of event times with censored total time. The presmoothed estimators $p_{ij}^{prs}(s,t)$ are step functions, with jumps at the observed (censored or uncensored) times. In this aspect they differ from landmark estimators ($p_{ij}^{ldm}(s,t)$) whose jumps are restricted to the uncensored times. In the limit case of no presmoothing, the Presmoothed Landmark estimator reduces to the landmark estimator.

In practice, estimation of the variance is needed for inference purposes. To this end, resampling techniques such as the bootstrap can be used. These methods can be used to construct confidence limits based on the bootstrap (e.g., using the basic or the percentile method) and thus to confirm if the proposed methods lead to a reduction in the variability of the estimators proposed in this section. These resampling techniques can be easily implemented using the R package described in Section 5.

Simulations reported in Section 3 reveal that the proposed estimators are virtually unbiased and that they may achieve good efficiency levels when compared to the unsmoothed landmark estimators.

2.4. Including covariates

In this section, we will explain how to introduce covariate information in the unsmoothed landmark estimators, $\widehat{p}_{ij}^{ldm}(s,t)$. In particular, we are interested in estimating the conditional transition probabilities $p_{ij}(s,t|X = x)$ that can be computed for any times s and t, $s < t$, but conditional to a given continuous covariate X that could either be a baseline covariate or a current covariate that is observed for an individual before the individual makes a particular transition of interest. Discrete covariates can be also included by splitting the sample for each level of the covariate and repeating the described procedures for each subsample.

To account for the covariate effect, one standard method is to consider estimators based on a Cox's model (Cox, 1972), with the corresponding baseline hazard function estimated by the Breslow's method (Breslow, 1972). Flexible effects of the covariates on the transition probabilities as those depicted in Figure 5 can be obtained using an alternative approach which introduces local smoothing by means of kernel weights based on local constant (Nadaraya-Watson) regression (Nadaraya 1965; Watson 1964).

Nonparametric estimators of the conditional transition probabilities have been recently proposed by Meira-Machado et al. (2015). These authors propose to estimate $p_{ij}(s,t \mid X=x)$ via estimation of the conditional expectations such as $E[\psi(Z,T) \mid X=x]$, where ψ is a general function defined over Z and T. Following the ideas described in Meira-Machado et al. (2015), the conditional transition probabilities are defined as follows:

$$\begin{aligned} p_{11}(s,t \mid X=x) &= \frac{1-P(Z \leq t \mid X=x)}{1-P(Z \leq s \mid X=x)}, \\ p_{13}(s,t \mid X=x) &= \frac{P(Z>s, T \leq t \mid X=x)}{1-P(Z \leq s \mid X=x)} \\ p_{23}(s,t \mid X=x) &= \frac{P(Z \leq s, s<T \leq t \mid X=x)}{P(Z \leq s \mid X=x)-P(T \leq s \mid X=x)}. \end{aligned} \tag{8}$$

The conditional transition probability $p_{11}(s,t \mid X=x)$, the denominator of $p_{13}(s,t \mid X=x)$ and the denominator of $p_{23}(s,t \mid X=x)$ involve the estimation of the conditional distribution/survival function of the response, given the covariate under random right censoring. This topic was introduced by Beran (1981) and was further studied by several authors (see e.g. papers by Akritas, 1994; van Keilegom et al., 2001; Akritas and van Keilegom, 2003). Their proposals can be used to estimate for instances the conditional distribution function of $Z \mid X=x$, that is, $F_{X=x}(u)=P(Z \leq u \mid X=x)$ which we denote by $\widehat{F}_{X=x}$ or simply by $\widehat{F}_x$. This can be done using the estimator introduced by Beran (1981),

$$\widehat{F}_{X=x}(u) = 1 - \prod_{\widetilde{Z}_i \leq u, \Delta_{1i}=1} \left[1 - \frac{NW(x,X_i,h)}{\sum_{j=1}^{n} I(\widetilde{Z}_j \geq \widetilde{Z}_i) NW(x,X_j,h)} \right], \tag{9}$$

where $NW(x,X_i,h)$ are the Nadaraya-Watson (NW) weights (Nadaraya, 1965; Watson, 1964)

$$NW(x,X_i,h) = \frac{D((x-X_i)/h)}{\sum_{j=1}^{n} D((x-X_j)/h)}$$

where D is a known probability density function (the kernel function) and h is a bandwidth.

The remaining quantities in the computation of the conditional transition probabilities involve conditional expectations of particular transformations of the pair (Z,T) given X, $E[\psi(Z,T) \mid X = x]$ which can not be estimated so simply.

In the absence of censoring, to estimate the conditional expectations $E[\psi(Z,T) \mid X = x]$ we may use kernel smoothing techniques by calculating a local average of the $\psi(Z,T)$, that is, as follows:

$$\widehat{E}[\psi(Z,T) \mid X = x] = \sum_{i=1}^{n} NW(x,X_i,h)\psi(\widetilde{Z}_i,\widetilde{T}_i),$$

where $NW(x,X_i,h)$ are the Nadaraya-Watson (NW) weights.

To handle right censoring Meira-Machado et al. (2015) propose the use of inverse of probability censoring weighting. Assuming that $\forall x$, $P(\widetilde{Z} > s, \widetilde{T} \leq t \mid X = x) > 0$ and $P(\widetilde{Z} \leq s, s < \widetilde{T} \leq t \mid X = x) > 0$, we have the following:

$$E[I(Z > s, T \leq t) \mid X = x] = E[I(\widetilde{Z} > s, s < \widetilde{T} \leq t)\Delta / K_X(\widetilde{T}) \mid X = x],$$

$$E[I(Z \leq s, T > s) \mid X = x] = E[I(\widetilde{Z} \leq s, \widetilde{T} \leq t)\Delta / K_X(\widetilde{T}) \mid X = x]$$

where K_X denotes the conditional survival function of the censoring variable C given the covariate X, that is $K_{X=x}(u) = P(C > u \mid X = x)$. Let $\widehat{K}_{X=x}$ denote Beran's estimator of K_X. Based on this, the following nonparametric estimators of the conditional transition probabilities can be introduced:

$$\widehat{p}_{11}(s,t \mid X = x) = \frac{1 - \widehat{F}_x(t)}{1 - \widehat{F}_x(s)}, \tag{10}$$

$$\widehat{p}_{13}(s,t \mid X = x) = \frac{1}{1 - \widehat{F}_x(s)} \sum_{i=1}^{n} \frac{NW(x,X_i,h_1) I(\widetilde{Z}_i > s, \widetilde{T}_i \leq t)\Delta_i}{\widehat{K}_{X_i}(\widetilde{T}_i)}, \tag{11}$$

and

$$\widehat{p}_{23}(s,t \mid X = x) = \frac{\sum_{i=1}^{n} NW(x,X_i,h_1) I(\widetilde{Z}_i \leq s, s < \widetilde{T}_i \leq t)\Delta_i / \widehat{K}_{X_i}(\widetilde{T}_i)}{\widehat{F}_x(s) - \widehat{H}_x(s)}, \tag{12}$$

where $\widehat{H}_x$ denote Beran's estimator of the conditional distribution of $T \mid X = x$.

Similar ideas as those explained above can be used to introduce nonparametric estimators for the conditional transition probabilities based on landmark. For example, given the time point s, the estimation of the conditional transition probabilities $p_{11}(s,t \mid X = x)$ and $p_{13}(s,t \mid X = x)$ are restricted to the individuals in State 1 at time s. Thus, the landmark estimators for these quantities are given as follows:

$$\widetilde{p}_{11}(s,t \mid X=x) = 1 - \sum_{i=1}^{n_{1s}} NW(x, X_{(i)}^{(s)}, h_1) \frac{I(\widetilde{Z}_{(i)}^{(s)} \leq t)\Delta_{1[i]}^{(s)}}{\widehat{G}_{X_i}(\widetilde{Z}_{(i)}^{(s)})}, \tag{13}$$

and

$$\widetilde{p}_{13}(s,t \mid X=x) = \sum_{i=1}^{n_{1s}} NW(x, X_{(i)}^{(s)}, h_1) \frac{I(\widetilde{T}_{(i)}^{(s)} \leq t)\Delta_{[i]}^{(s)}}{\widehat{K}_{1,X_i}(\widetilde{T}_{(i)}^{(s)})}, \tag{14}$$

where $\widehat{G}_X$ and $\widehat{K}_{1,X}$ are Beran's estimators for the conditional survival function of the censoring variable of the sojourn time in State 1 (respectively, total time) given X in subset $\mathscr{S}_1$.

Similarly, the built of the landmark estimator of the conditional transition probability $p_{23}(s,t \mid X=x)$ is restricted to the individuals in State 2 at time s:

$$\widetilde{p}_{23}(s,t \mid X=x) = \sum_{i=1}^{n_{2s}} NW(x, X_{(i)}^{[s]}, h_2) \frac{I(\widetilde{T}_{(i)}^{[s]} \leq t)\Delta_{[i]}^{[s]}}{\widehat{K}_{2,X_i}(\widetilde{T}_{(i)}^{[s]})}, \tag{15}$$

where $\widehat{K}_{2,X}$ is Beran's estimator of the conditional survival function of the censoring variable of the total time given X in subset $\mathscr{S}_2$.

Simulation results (not reported here) reveal that the landmark based estimators $\widetilde{p}_{12}(s,t \mid X=x)$ and $\widetilde{p}_{13}(s,t \mid X=x)$ perform favourably when compared to $\widehat{p}_{12}(s,t \mid X=x)$ and $\widehat{p}_{13}(s,t \mid X=x)$, respectively. In contrast, the landmark estimator $\widetilde{p}_{23}(s,t \mid X=x)$ have a worst performance when compared to $\widehat{p}_{23}(s,t \mid X=x)$ particularly when computed at time points s for which few individuals are observed in State 2.

3. Simulation study

In this section, we report the results of a simulation study carried out to investigate the empirical behaviour of the estimators, introduced in Section 2, for finite sample sizes. More specifically, the landmark unsmoothed estimators, $\widetilde{p}_{ij}^{ldm}(s,t)$, with the smoothed estimators, $\widetilde{p}_{ij}^{crs}(s,t)$, based on cubic regression splines and the semiparametric presmoothed estimators, $\widetilde{p}_{ij}^{prs}(s,t)$.

To simulate the data in the illness-death model, we use the same scenario as that described in Amorim et al. (2011) and de Uña-Álvarez and Meira-Machado (2015). We separately consider the subjects passing through State 2 at some time, and those who directly go to the absorbing State 3. For the first subgroup of individuals ($\rho = 1$), the successive gap times $(Z, T-Z)$ are simulated according to the bivariate exponential distribution

$$F_{12}(u,v) = F_1(u)F_2(v)\left[1+\gamma\left\{1-F_1(u)\right\}\left\{1-F_2(v)\right\}\right]$$

with exponential marginal distribution functions with rate parameter 1. The single parameter γ controls the amount of dependency between the gap times. The parameter γ was set to 0 for simulating independent gap times, and also to 1, corresponding to 0.25 correlation between Z and $T-Z$. The simulation procedure is as follows:

Step 1. Draw $\rho \sim Ber(p)$ where p is the proportion of subjects passing through State 2.

Step 2. If $\rho = 1$ then:

(2.1) $V_1 \sim U(0,1), V_2 \sim U(0,1)$ are independently generated;

(2.2) $U_1 = V_1, A = \gamma(2U_1-1)-1, B = (1-\gamma(2U_1-1))^2 + 4\gamma V_2(2U_1-1)$

(2.3) $U_2 = 2V_2/\left(\sqrt{B}-A\right)$

(2.4) $Z = \ln(1/(1-U_1)), T = \ln(1/(1-U_2)) + Z$

If $\rho = 0$ then $Z = \ln(1/(1-U(0,1)))$.

Situations with $p = 1$ corresponds to the three-state progressive model, in which a direct transition to State 3 is not allowed. In our simulation we consider $p = 0.7$. An independent uniform censoring time C is generated, according to models $U[0,4]$ and $U[0,3]$. The first model results in 24% of censoring on the first gap time Z, and in 47% of censoring on the second gap time $T-Z$, for those individuals with $\rho = 1$. The second model increases these censoring levels to 32% and about 57%, respectively.

For each simulated setting we derived the analytic expression of $p_{ij}(s,t)$ for six different points (s,t) $(s<t)$, corresponding to combinations of the percentiles 20%, 40%, 60% and 80% of the marginal distributions of the gap times. Sample sizes of 100, 250 and 500 were considered. In each simulation, 1000 samples were generated and for each of the three estimators we obtain the mean bias, the standard deviation (SD), and the mean square error (MSE) based on the 1000 Monte Carlo replicates. Table 1 reports the results for the transition probabilities $p_{12}(s,t)$ and $p_{23}(s,t)$ for the case with dependent gap times; the results for independent gap times (not shown) are similar.

As would be expected, results reported in Table 1 reveal that the performance of all methods is poorer at the right tail (i.e., larger values of s and t) where the censoring effects are stronger. At these points the SD is in most cases higher. The SD decreases with an increase in the sample size and with a decrease of the censoring percentage. All methods proposed in this article obtain in all settings a small bias.

Results reported in Table 1 reveal that the SD clearly dominates the performance of the proposed estimators in most cases. This is particularly clear when comparing the semiparametric estimators with the unsmoothed landmark estimators. The semiparamet-

Table 1: *Bias and standard deviation (SD) for the three estimators of $p_{ij}(s,t)$. The MSE of $\widehat{p}_{ij}^{prs}(s,t)$ and $\widehat{p}_{ij}^{crs}(s,t)$ relative to $\widehat{p}_{ij}^{ldm}(s,t)$ are also given. Scenario of correlated exponential gap times with three sample sizes and two censoring levels.*

		$\widehat{p}_{12}^{ldm}(s,t)$		$\widehat{p}_{12}^{prs}(s,t)$		$\widehat{p}_{12}^{crs}(s,t)$			
		bias	SD	bias	SD	bias	SD	MSE^{ldm}/MSE^{prs}	MSE^{ldm}/MSE^{crs}
$(s,t)=$	(.2231,.5108)								
$n=100$	$C\sim U[0,4]$	−0.0017	0.0408	−0.0013	0.0368	0.0046	0.0377	1.2086	1.1532
	$C\sim U[0,3]$	−0.0017	0.0415	−0.0011	0.0375	0.0026	0.0385	1.2279	1.1586
$n=250$	$C\sim U[0,4]$	0.0006	0.0263	0.0015	0.0235	0.0040	0.0249	1.2468	1.0882
	$C\sim U[0,3]$	0.0006	0.0268	0.0017	0.0240	0.0022	0.0250	1.2388	1.1410
$n=500$	$C\sim U[0,4]$	0.0017	0.0188	0.0024	0.0170	0.0039	0.0170	1.2075	1.1718
	$C\sim U[0,3]$	0.0016	0.0189	0.0027	0.0169	0.0022	0.0171	1.2325	1.2100
$(s,t)=$	(.2231,.9163)								
$n=100$	$C\sim U[0,4]$	−0.0028	0.0518	−0.0025	0.0473	0.0017	0.0497	1.2034	1.0881
	$C\sim U[0,3]$	−0.0030	0.0530	−0.0035	0.0487	0.0003	0.0525	1.1845	1.0221
$n=250$	$C\sim U[0,4]$	<0.0001	0.0338	0.0027	0.0311	0.0006	0.0320	1.1855	1.1153
	$C\sim U[0,3]$	−0.0001	0.0352	−0.0007	0.0320	0.0007	0.0328	1.2078	1.1512
$n=500$	$C\sim U[0,4]$	0.0010	0.0235	0.0008	0.0215	0.0002	0.0221	1.1970	1.1326
	$C\sim U[0,3]$	0.0007	0.0243	−0.0003	0.0219	0.0017	0.0239	1.2333	1.0295
$(s,t)=$	(.5108,1.6094)								
$n=100$	$C\sim U[0,4]$	0.0051	0.0704	0.0025	0.0642	0.0024	0.0691	1.2084	1.0422
	$C\sim U[0,3]$	0.0047	0.0780	−0.0004	0.0695	0.0029	0.0774	1.2647	1.0175
$n=250$	$C\sim U[0,4]$	0.0022	0.0438	−0.0007	0.0397	0.0011	0.0444	1.2176	0.9763
	$C\sim U[0,3]$	0.0019	0.0489	−0.0028	0.0435	0.0007	0.0487	1.2607	1.0097
$n=500$	$C\sim U[0,4]$	0.0005	0.0301	−0.0015	0.0273	0.0015	0.0301	1.2112	0.9979
	$C\sim U[0,3]$	0.0008	0.0337	−0.0036	0.0296	0.0015	0.0333	1.2745	1.0227
$(s,t)=$	(.9163,1.6094)								
$n=100$	$C\sim U[0,4]$	0.0055	0.0848	0.0022	0.0775	0.0007	0.0883	1.2015	0.9261
	$C\sim U[0,3]$	0.0053	0.0956	0.0008	0.0873	−0.0070	0.0978	1.2019	0.9537
$n=250$	$C\sim U[0,4]$	0.0029	0.0547	−0.0005	0.0492	0.0027	0.0543	1.2362	1.0152
	$C\sim U[0,3]$	0.0026	0.0610	−0.0020	0.0539	0.0021	0.0589	1.2820	1.0733
$n=500$	$C\sim U[0,4]$	<0.0001	0.0383	−0.0032	0.0346	0.0022	0.0367	1.2123	1.0862
	$C\sim U[0,3]$	0.0008	0.0417	−0.0040	0.0371	0.0014	0.0408	1.2491	1.0440
		$\widehat{p}_{23}^{ldm}(s,t)$		$\widehat{p}_{23}^{prs}(s,t)$		$\widehat{p}_{23}^{crs}(s,t)$			
$(s,t)=$	(.2231,.5108)								
$n=100$	$C\sim U[0,4]$	0.0016	0.1687	−0.0028	0.1693	0.0042	0.1475	0.9937	1.3054
	$C\sim U[0,3]$	0.0024	0.1722	−0.0006	0.1706	0.0040	0.1535	1.0203	1.2570
$n=250$	$C\sim U[0,4]$	−0.0068	0.0967	−0.0087	0.0946	0.0028	0.0965	1.0422	1.0088
	$C\sim U[0,3]$	−0.0068	0.0971	−0.0063	0.0957	0.0020	0.0974	1.0294	0.9983
$n=500$	$C\sim U[0,4]$	−0.0017	0.0692	−0.0014	0.0677	0.0033	0.0661	1.0463	1.0932
	$C\sim U[0,3]$	−0.0015	0.0704	<0.0001	0.0685	0.0025	0.0658	1.0559	1.1409
$(s,t)=$	(.2231,.9163)								
$n=100$	$C\sim U[0,4]$	0.0015	0.1615	−0.0003	0.1566	<0.0001	0.1456	1.0633	1.2259
	$C\sim U[0,3]$	0.0009	0.1671	−0.0005	0.1579	−0.0034	0.1550	1.1205	1.1612
$n=250$	$C\sim U[0,4]$	0.0021	0.0939	0.0018	0.0921	0.0003	0.0910	1.0405	1.0653
	$C\sim U[0,3]$	0.0024	0.0972	0.0008	0.0943	0.0023	0.0962	1.0632	1.0209
$n=500$	$C\sim U[0,4]$	0.0031	0.0657	0.0021	0.0647	<0.0001	0.0626	1.0323	1.1039
	$C\sim U[0,3]$	0.0033	0.0691	0.0021	0.0666	0.0001	0.0651	1.0794	1.1282
$(s,t)=$	(.5108,1.6094)								
$n=100$	$C\sim U[0,4]$	−0.0006	0.1247	−0.0053	0.1169	0.0019	0.1229	1.1352	1.0293
	$C\sim U[0,3]$	−0.0058	0.1358	−0.0137	0.1268	0.0020	0.1329	1.1344	1.0456
$n=250$	$C\sim U[0,4]$	0.0005	0.0768	−0.0013	0.0731	0.0012	0.0738	1.1035	1.0828
	$C\sim U[0,3]$	0.0020	0.0835	−0.0068	0.0779	0.0012	0.0807	1.1388	1.0702
$n=500$	$C\sim U[0,4]$	0.0019	0.0540	0.0006	0.0517	0.0007	0.0522	1.0902	1.0711
	$C\sim U[0,3]$	0.0006	0.0604	−0.0038	0.0563	0.0001	0.0573	1.1445	1.1110
$(s,t)=$	(.9163,1.6094)								
$n=100$	$C\sim U[0,4]$	−0.0085	0.1391	−0.0111	0.1335	0.0023	0.1388	1.0816	1.0078
	$C\sim U[0,3]$	−0.0086	0.1525	−0.0122	0.1422	0.0070	0.1460	1.1448	1.0920
$n=250$	$C\sim U[0,4]$	−0.0031	0.0870	−0.0040	0.0828	0.0024	0.0836	1.1041	1.0841
	$C\sim U[0,3]$	−0.0042	0.0979	−0.0057	0.0904	0.0019	0.0922	1.1690	1.1290
$n=500$	$C\sim U[0,4]$	−0.0009	0.0593	−0.0004	0.0564	−0.0006	0.0590	1.1056	1.0121
	$C\sim U[0,3]$	−0.0022	0.0665	−0.0034	0.0616	−0.0012	0.0647	1.1635	1.0572

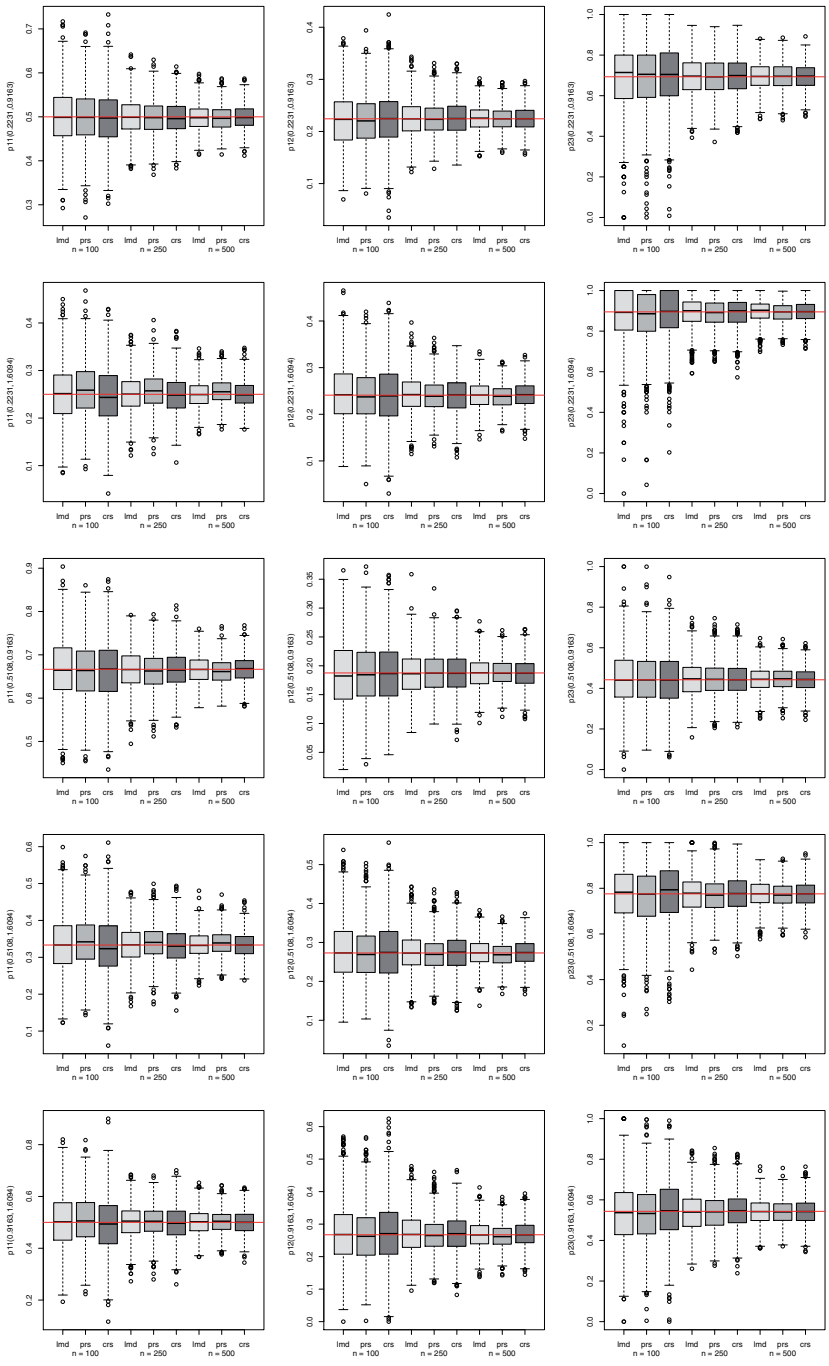

Figure 2: *Boxplots of the M = 1000 estimates of the transition probabilities of the three estimators, with three different sample sizes and correlated exponential gap times. Censoring times were generated from an uniform distribution on* $[0,3]$. *Horizontal solid red line corresponds to the true value of the transition probability.*

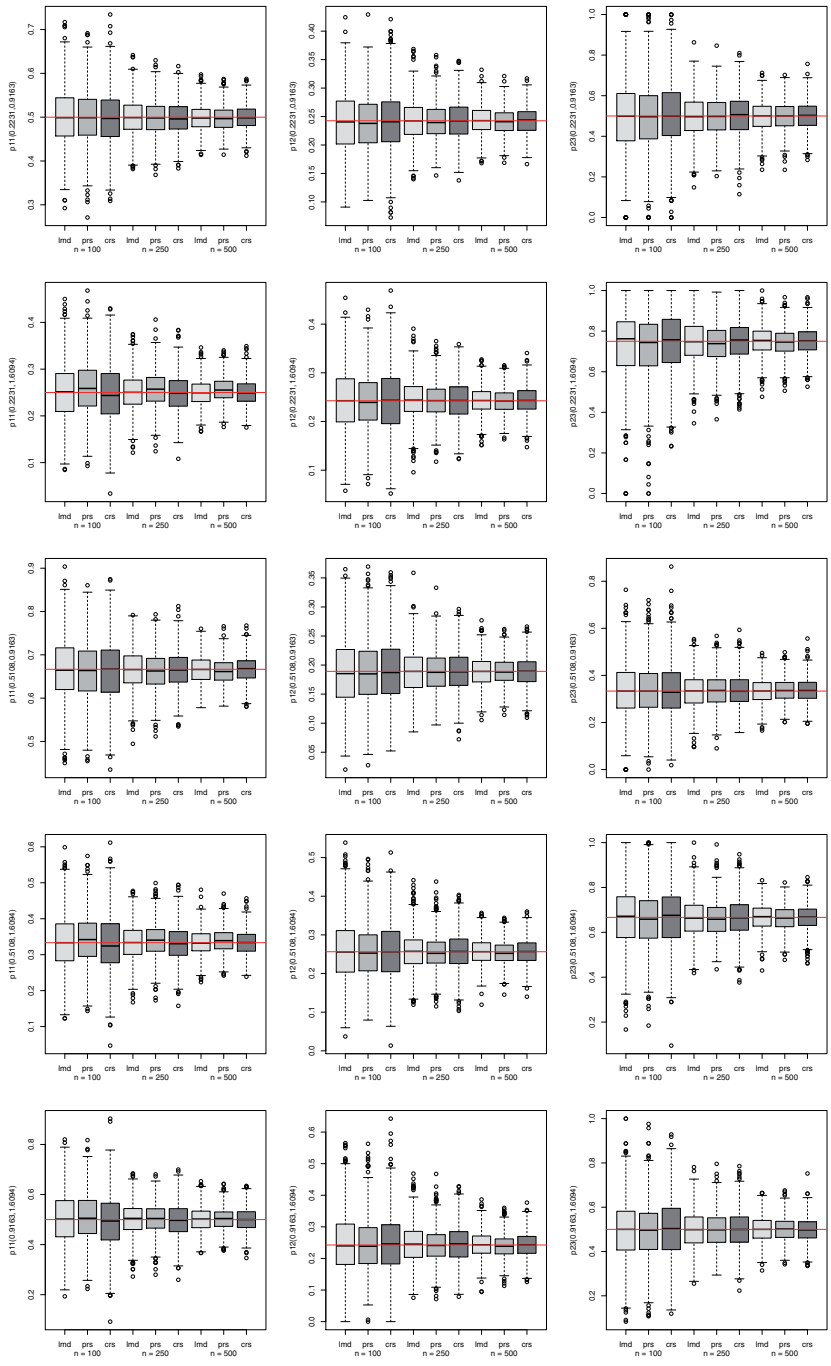

Figure 3: *Boxplots of the $M = 1000$ estimates of the transition probabilities of the three estimators, with three different sample sizes and uncorrelated exponential gap times. Censoring times were generated from an uniform distribution on $[0,3]$. Horizontal solid red line corresponds to the true value of the transition probability.*

ric estimator achieve better results with less SD and less MSE. This can be seen by the relative efficiency between the semiparametric estimator and the unsmoothed landmark estimator that was measured by the ratio between their corresponding MSEs. The semiparametric estimators reported a smaller MSE in most cases. It can also be observed that the relative advantages of the semiparametric estimator is greater for higher censoring percentages. This advantage is also apparently greater when estimating the transition probability $p_{12}(s,t)$.

For completeness purposes we show in Figures 2 and 3 the boxplots of the estimates of the transition probabilities based on the 1000 Monte Carlo replicates for the three estimators, with different sample sizes, for correlated exponential gap times (Figure 2) and independent gap times (Figure 3). Plots shown in these figures were obtained for the higher censoring levels ($C \sim U[0,3]$). In addition to the transition probabilities reported in Table 1 this figure also show the performance of the transition probability $p_{11}(s,t)$ for all methods. The boxplots shown in Figures 2 and 3 reveal some results which are in agree with our findings reported in Table 1. From these plots it can be seen that all methods have small bias and confirm the less variability of the semiparametric estimators.

Despite of offering a small bias, the bias associated to presmoothed estimators is in general larger than that of the unsmoothed landmark estimator. This bias component may be larger when there is some miss-specification in the chosen parametric model. Importantly, the validity of a given model for presmoothing can be checked graphically or formally, by applying a goodness-of-fit tests (e.g. Hosmer and Lemeshow (1989)). This implies that the risk of introducing a large bias through a miss-specified model can be controlled in practice.

4. German breast cancer study

Breast cancer is one of the most commonly diagnosed cancers in women. Prognosis of this carcinoma is related to a large variety of clinical and pathological factors such as age, tumor size, histological grade, lymph node involvement, and hormone receptor status. Another significant prognostic factor for these patients in overall survival is the presence of a recurrence. Traditionally, the effect of these time-dependent covariates is studied using extensions of the Cox proportional hazards model (Cox, 1972; Genser and Wernecke, 2005). The analysis of such studies can also be successfully performed using a multi-state model (Pérez-Ocón et al., 2001; Putter et al., 2007; Meira-Machado et al., 2009).

Several studies have been developed over the last decades regarding breast cancer. Between 1983 and 1989, four clinical trials were conducted by the German Breast Cancer Study Group (GBSG) including 2746 patients with primary node positive breast cancer. Details about these studies can be found in the paper by Schumacher et al. (1994).

Among other papers, these data were used by Schmoor et al. (2000) and Meier-Hirmer and Schumacher (2013). In both cases the main goal was to evaluate the effect on future prognosis of an isolated locoregional recurrence (ILRR). While Schmoor et al. (2000) used a Cox proportional hazards model, Meier-Hirmer and Schumacher (2013) used an illness-death model to investigate the influence of the time-dependent covariate 'recurrence'. Both studies conclude, among other things, that the increased risk after ILRR decreased significantly with increasing time since ILRR. In this paper we use data from the second trial in which a total of 720 women with primary node positive breast cancer is recruited in the period between July 1984 and December 1989. The data is available at the University of Massachusetts website for statistical software information as well as part of the R packages mfp, TH.data and flexsurv. The data which was also used by Sauerbrei and Royston (1999) considers 686 patients who had complete data for the two event times (time to recurrence and time to death). In this study, patients were followed from the date of breast cancer diagnosis until censoring or dying from breast cancer. From the total of 686 women, 299 developed a recurrence and 171 died. Besides the two event times and the corresponding indicator statuses a vector of covariates including age at acceptance tumor size, number of positive lymph nodes, progesterone and estrogen receptor status, menopausal status and tumor grade are also available. The covariate 'recurrence' is the only time-dependent covariate, while the other covariates included are fixed. This covariate can be considered as an intermediate transient state and modeled using an illness-death model with states 'Alive and disease-free', 'Alive with Recurrence' and 'Dead'. In this section, we present plots for the three different methods to estimate the transition probabilities described in Section 2. Figure 4 reports estimated transition probabilities for $p_{11}(s,t)$, $p_{12}(s,t)$ and $p_{23}(s,t)$, for fixed values $s = 365$ and $s = 730$ (days), along time t (corresponding to 1 and 2 years after surgery). Plots shown in these figure also show the pointwise bootstrap confidence bands of the unsmoothed method. Estimators for all three methods shown in these plots report roughly the same estimates. Minor differences are appreciated when comparing the nonparametric unsmoothed method with their counterparts (the semiparametric presmoothed approach and the method based on cubic regression splines) which is in agree with our findings in the simulation study.

Plots shown at the top of Figure 4 provide the probabilities of being alive and without recurrence for the individuals who are disease free 1 year (Figure 4, top left) and 2 years (Figure 4, top right) after surgery (i.e. $p_{11}(s,t)$). These are monotonous decreasing curves. The curve do not decrease to zero due to a (disease free) censoring rate of about 56.4% (387 woman remain alive and disease free until the end of study). In addition, one can observe that these probabilities increase with an increase of the value of s. Similar conclusions can be obtained from the plots shown at the bottom of Figure 4, in which the transition probability $p_{23}(s,t)$ is estimated through the three methods. These plots report one minus the survival fraction along time, among the individuals in the recurrence state 1 year (Figure 4, bottom left) and 2 years (Figure 4, bottom right) after surgery. It can be observed from these plots that the survivorship is smaller for the first

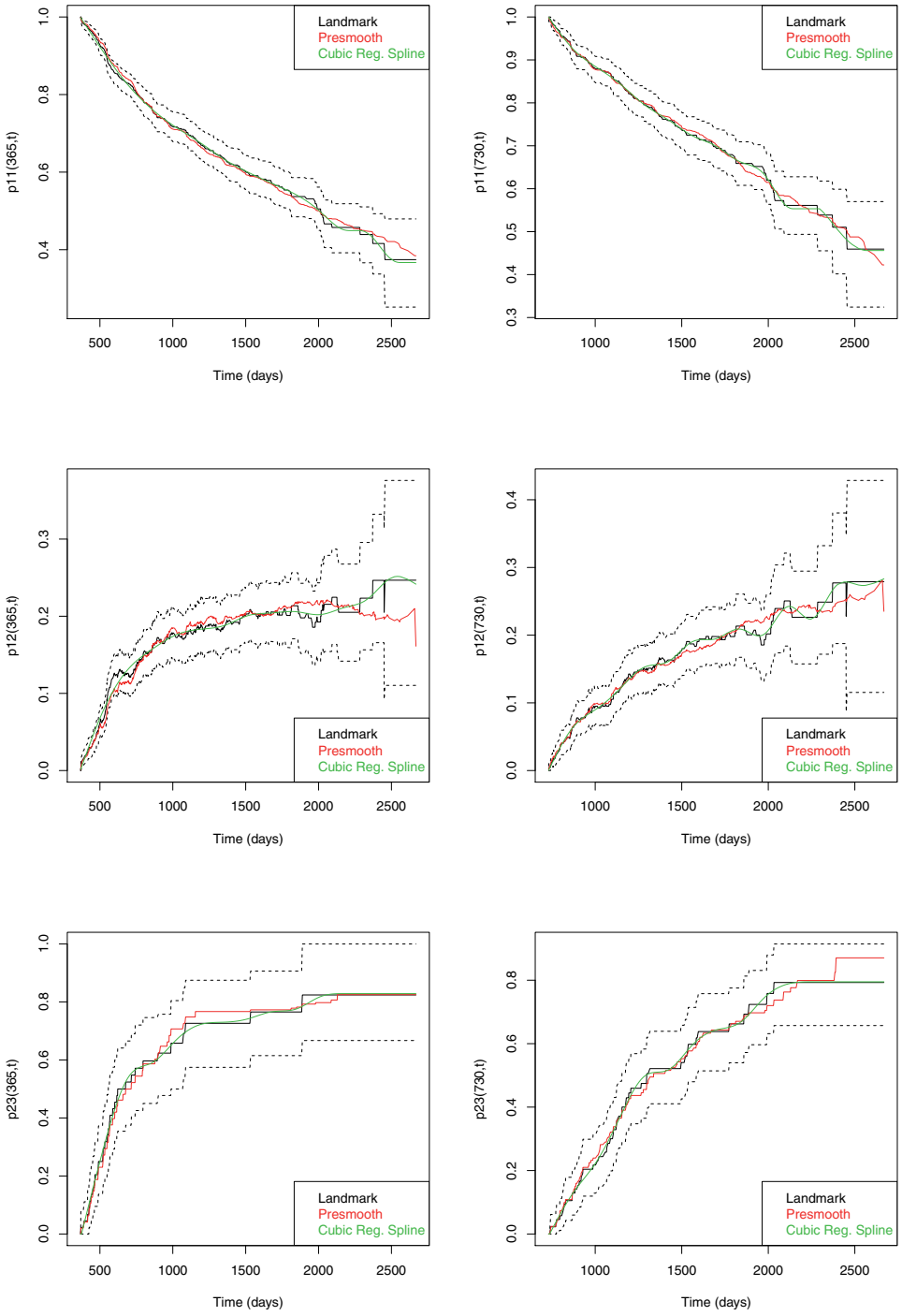

Figure 4: *Estimates of the transition probabilities $p_{11}(s,t)$, $p_{12}(s,t)$ and $p_{23}(s,t)$ for $s = 365$ (left) and $s = 730$ (right) using the three methods (landmark, presmoothing and cubic regression splines). Pointwise confidence intervals of the landmark method is also shown. Breast cancer data.*

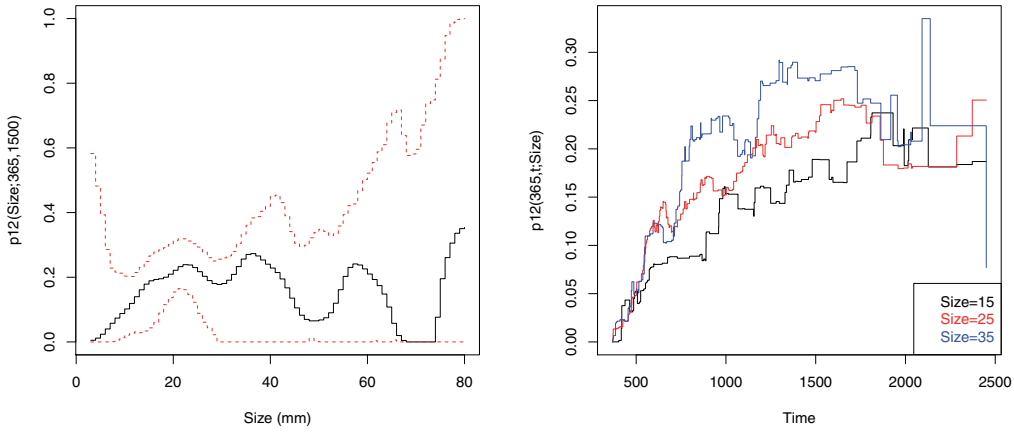

Figure 5: *Landmark estimates of the conditional transition probabilities* $p_{12}(s=365,t=1500 \mid size)$ *(left) and* $p_{12}(365,t \mid size)$ *(right). Breast cancer data.*

group for lower values of t, suggesting a negative impact of an earlier recurrence time. In contrast to these plots, curves for the transition probability $p_{12}(s,t)$ are not necessarily monotonous. Plots shown for this transition probability allows for an inspection along time of the probability of being alive with recurrence for the individuals who are disease free at 1 and 2 years after surgery. Since the recurrence state is transient, in general, this curve is first increasing and then decreasing. However, in this case, for $s = 365$, the curve has a rapid increase at lower times and afterwards remain roughly constant. The increase is more gradual for $s = 730$. The reason why the curve does not decreases can be explained by the percentage of about 46% of woman that remain in the recurrence state at the end of study. Departures between estimated curves can be more appreciated for larger time values where the censoring effects are stronger.

Figure 5 depict the landmark estimates of the conditional transition probability on the recurrence transition. Plot shown at the left depicts the estimates of the transition probability $p_{12}(\text{size};365,1500)$ as functions of the covariate tumor size together with a 95% pointwise confidence bands based on simple bootstrap which resamples each datum with probability $1/n$. Plot at the right depicts the estimates of the transition probability $p_{12}(365,t)$ conditional on the covariate tumor size. The effects of tumor size according to three groups depicted in these plots, which are purely nonparametric, indicate the real influence of this covariate in the recurrence transition. Both plots are in agreement and indicate that patients with higher tumor sizes have a larger probability of recurrence. To compute the conditional transition probabilities shown in this figure we have used a common bandwidth selector and Gaussian kernels. To this end we have used the dpik function which is available from the R KernSmooth package.

5. Software development

To provide the biomedical researchers with an easy-to-use tool for obtaining estimates of the transition probabilities we develop an R package called tprob. This package can be used to implement all nonparametric and semiparametric estimators for the transition probabilities discussed in Section 2. In addition, estimators are also implemented that account for the influence of covariates. Bootstrap confidence bands are provided for all methods. This package is composed by several functions that allow users to obtain estimates and plots of the transition probabilities. Details on the usage of these functions can be obtained with the corresponding help pages. The CPU time needed for running some of the proposed methodologies varies according to whether bootstrap confidence bands are requested or not, the sample size, and the type of processor in the computer. To minimize these problems the most computationally demanding parts of the code were developed and implemented in the C programming language. This software is available at the author web site `http://w3.math.uminho.pt/~lmachado/R/tprob`.

6. Discussion

There have been several recent contributions for the estimation of the transition probabilities in the context of non-Markov multi-state models. Meira-Machado et al. (2006) introduced a substitute for the Aalen-Johansen estimator in the case of a non-Markov illness-death model. They showed that the new estimator may behave much more efficiently than the Aalen-Johansen when the Markov assumption does not hold. However, the proposal of Meira-Machado et al. (2006) has the drawback of requiring that the support of the censoring distribution contains the support of the lifetime distribution, otherwise they only report valid estimators for truncated transition probabilities. Recently, the problem of estimating the transition probabilities in a non-Markov illness-death model has been reviewed, and new estimators have been proposed which are consistent regardless the Markov condition and the referred assumption on the censoring support. These estimators are built by considering specific subsets of individuals (namely, those observed to be in a given state at a prespecified time point s for which the ordinary Kaplan-Meier survival function leads to a consistent estimator of the target. As a weakness, it provides large standard errors for large values of s and higher censoring percentages.

In this article we propose two approaches that can be used to reduce the variability of the proposed estimator. A simple approach is based on spline smoothing (cubic regression splines). Another valid approach is to consider a semiparametric estimator based on a presmoothed version of the Kaplan-Meier estimator. The provided simulations suggest that both approaches are preferable to the original nonparametric estimator, since they often have less variance while providing more reliable curves. Between the two new methods, the one based on presmoothing is recommended.

Acknowledgements

This research was financed by FEDER Funds through "Programa Operacional Factores de Competitividade-COMPETE" and by Portuguese Funds through FCT-"Fundação para a Ciência e a Tecnologia", in the form of grant UID/MAT/00013/2013. We thank the two anonymous referees for comments and suggestions which have improved the presentation of the paper.

References

Aalen, O. and Johansen, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, 5, 141–150.

Akritas, M. (1994). Nearest neighbor estimation of a bivariate distribution under random censoring. *The Annals of Statistics*, 22, 1299–1327.

Akritas, M.G. and van Keilegom, I. (2003). Estimation of bivariate and marginal distributions with censored data, *Journal of Royal Statistical Society, B*, 65, 457–471.

Allignol, A., Beyersmann, J., Gerds, T. and Latouche, A. (2014). A competing risks approach for nonparametric estimation of transition probabilities in an non-Markov illness-death model. *Lifetime Data Analysis*, 20, 495–513.

Amorim, A., de Uña-Álvarez, J. and Meira-Machado, L. (2011). Presmoothing the transition probabilities in the illness-death model. *Statistics & Probability Letters*, 81, 797–806.

Andersen, P.K. and Keiding, N. (2002). Multi-state models for event history analysis. *Statistical Methods Medical Research*, 11, 91–115.

Beran, R. (1981). *Nonparametric Regression with Randomly Censored Survival Data*. Technical report, Univ. California, Berkeley.

Breslow, N.E. (1972). Discussion of the paper by D. R. Cox. *Journal of Royal Statistical Society, B*, 34, 216–217.

Cao, R. and Jácome, M.A. (2004). Presmoothed kernel density estimator for censored data, *Journal of Nonparametric Statistics*, 16, 289–309.

Cao, R., López-de-Ullibarri, I., Janssen, P. and Veraverbeke N (2005). Presmoothed Kaplan-Meier and Nelson-Aalen estimators, *Journal of Nonparametric Statistics*, 17, 31–56.

Cox, D.R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34, 187–200.

de Uña-Álvarez, J. and Rodríguez-Campos, C. (2004). Strong consistency of presmoothed Kaplan-Meier integrals when covariables are present. *Statistics*, 38, 483–496.

de Uña-Álvarez, J. and Amorim, A.P. (2011). A semiparametric estimator of the bivariate distribution function for censored gap times. *Biometrical Journal*, 53, 113–127.

de Uña-Álvarez, J. and Meira-Machado, L. (2015). Nonparametric estimation of transition probabilities in the non-Markov illness-death model: a comparative study. *Biometrics*, 71, 364–375.

Dikta, G. (1998). On semiparametric random censorship models. *Journal of Statistical Planning and Inference*, 66, 253–279.

Durrleman, S. and Simon, R. (1989). Flexible regression models with cubic splines. *Statistics in Medicine*, 8, 551–561.

Genser, B. and Wernecke, K.D. (2005). Joint modelling of repeated transitions in follow-up data: a case study on breast cancer data. *Biometrical Journal*, 47, 388–401.

Gentleman, R.C., Lawless, F.F., Lindsey, J.C. and Yan, P. (1994) Multi-state Markov models for analysing incomplete disease history data with illustrations for HIV disease. *Statistics in Medicine*, 13, 805–821.

Hosmer, D.W. and Lemeshow, S. (1989). *Applied Logistic Regression.* New York: Wiley.

Jácome, M.A. and Iglesias, M.C. (2008). Presmoothed estimation with left truncated and right censored data. *Communications in Statistics/Theory and Methods*, 37, 2964–2983.

Jácome, M.A. and Iglesias, M.C. (2010). Presmoothed estimation of the density function with truncated and censored data. *Statistics*, 44, 217–234.

Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. em Journal of the American Statistical Association, 53, 457–481.

López-de-Ullibarri, I. and Jácome, M.A. (2013) survPresmooth: An R Package for Presmoothed Estimation in Survival Analysis. *Journal of Statistical Software*, 54, 1–26.

Meyer, M.C. (2008) Inference using shape-restricted regression splines, *The Annals of Applied Statistics*, 2, 1013–1033.

Meyer, M.C. (2012) Constrained penalized splines, *The Canadian Journal of Statistics*, 40, 190–206.

Meira-Machado, L., de Uña-Álvarez, J. and Cadarso-Suárez, C. (2006). Nonparametric estimation of transition probabilities in a non-Markov illness-death model. *Lifetime Data Analysis*, 12, 325–344.

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

Meira-Machado, L., de Uña-Álvarez, J. and Datta, S. (2015). Nonparametric estimation of conditional transition probabilities in a non-Markov illness-death model. *Computational Statistics*, 30, 377–397.

Meira-Machado, L., Sestelo, M. and Gonlçalves, A. (2016). Nonparametric estimation of the survival function for ordered multivariate failure time data: A comparative study. *Biometrical Journal*, 58, 623–634.

Meier-Hirmer, C. and Schumacher, M. (2013). Multi-state model for studying an intermediate event using time-dependent covariates: Application to breast cancer. *BMC Medical Research Methodology*, 13, 80.

Moreira, A.C., de Uña-Álvarez, J. and Meira-Machado, L. (2013). Presmoothing the Aalen-Johansen estimator in the illness-death model. *Electronic Journal of Statistics*, 7, 1491–1516.

Nadaraya, E. (1965). On nonparametric estimates of density functions and regression curves. Theory of Applied Probability, 10, 186–190.

Pérez-Ocón, R., Ruiz-Castro, J.E. and Gámiz-Pérez, M.L. (2001). Non-homogeneous Markov models in the analysis of survival after breast cancer. *Journal of the Royal Statistical Society: Series C*, 50, 111–124.

Putter, H, Fiocco, M. and Geskus, R.B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, 26, 2389–2430.

Pya, N. and Wood, S.N. (2015) Shape constrained additive models. *Statistical Computing*, 25, 543–559.

Satten, G.A. and Datta, S. (2001). The Kaplan-Meier Estimator as an inverse-probability-of-censoring weighted average. *American Statistician*, 55, 207–210.

Sauerbrei, W. and Royston, P. (1999). Building multivariable prognostic and diagnostic models: Transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society, A*, 161, 71–94.

Schmoor, C., Sauerbrei, W. Bastert, G., and Schumacher, M. (2000). Role of isolated locoregional recurrence of breast cancer: Results of four prospective studies. *Journal of Clinical Oncology*, 18, 1696–1708.

Schumacher, M., Bastert, G., Bojar, H., Hiibner, K., Olschewski, M., Sauerbrei, W., Schmoor, C., Beyerle, C., Neumann, R.L.A., and Rauschecker, H.F. for the German Breast Cancer Study Group (GBSG) (1994). A randomized 2 × 2 trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients. *Journal of Clinical Oncology*, 12, 2086–2093.

van Houwelingen, H.C. (2007). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*, 34, 70–85.

van Keilegom, I., Akritas, M. and Veraverbeke, N. (2001). Estimation of the conditional distribution in regression with censored data: A comparative study. *Computational Statistics and Data Analysis*, 35, 487–500.

Watson, G.S. (1964). Smooth regression analysis *Sankhya*, 26, 359–372.

Wood, S.N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC Press.