A comparison of presmoothing methods in the estimation of transition probabilities

Gustavo Soutinho\textsuperscript{a}, Luís Meira-Machado\textsuperscript{b}, and Pedro Oliveira\textsuperscript{c}

\textsuperscript{a}EPIUnit, Institute of Public Health of the University of Porto (ISPUP), Porto, Portugal; \textsuperscript{b}Dep. Mathematics and Applications, University of Minho, Guimaraes, Portugal; \textsuperscript{c}Population Studies, University of Porto, Porto, Portugal

ABSTRACT

One major goal in clinical applications of multi-state models is the estimation of transition probabilities. Estimators based on subsampling were recently introduced by de Uña-Alvarez and Meira-Machado 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. The idea behind the proposed estimators is to use a procedure based on (differences between) Kaplan-Meier estimators derived from a subset of the data consisting of all subjects observed to be in a given state at a given time. Subsampling, also referred to as landmarking, leads to small sample sizes and high percentage of censoring for which more efficient estimators are essential. Preliminary smoothing, also known as presmoothing, is a good alternative to these situations. The presmoothed estimators are obtained by replacing the censoring indicator variables in the classical definitions by values of a regression estimator. The behavior of the presmoothed landmark estimators for the transition probabilities is explored through simulation studies when different ways to estimate the conditional probability of uncensoring are used. Real data illustration is included.

1. Introduction

Multi-state models can be successfully used to model the movement of patients among a set of several states (Andersen et al. 1993; Hougaard 2000; Meira-Machado et al. 2009). These models are illustrated by a direct graph with rectangular boxes to represent possible states and with arrows between the states representing the allowed transitions. States can be transient or absorbing. Multi-state models have been widely used in medicine to investigate the progression of patients through different states of a disease such as cancer and AIDS. They have also been used in other areas of application including epidemiology, clinical trials, reliability studies in engineering, etc. Data in these applications are often subject to right censoring and possibly left truncation.

The simplest multi-state model is called the mortality model for survival data which consists of just two states (usually alive and death) and one transition. The competing
risks model (Andersen and Keiding, 2002; Putter, Fiocco, and Geskus 2007) can be seen as an extension of the simple mortality model for survival data in which each individual may ‘die’ due to any of several causes. A well-known and more complex multi-state model is the illness-death model. This model, also known as the disability model, can be used to study the incidence of the disease and the rate of death. In the irreversible version of this model, individuals may pass from the initial state (e.g., ‘alive and disease-free’; State 0), to the intermediate event or disease state (e.g., ‘diseased’; State 1) and then to the absorbing state (e.g., ‘dead’; State 2). Individuals are at risk of death in each transient state (states 0 and 1). One problem here may be to evaluate whether previously diseased subjects have the same rate of death as those who have been healthy all their lives. Unlike the survival or the competing risks model, the illness-death model is not necessarily Markovian, since the prognosis for an individual in the intermediate state may be influenced by the subject specific arrival time. Figure 1 shows the schematic diagram of transitions involved in the progressive illness-death model.

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 their illness process. The transition probabilities are a key quantity of interest since they allow for long-term predictions of the process. In case the process is Markovian, the transition probabilities depend on the current state (not on the history) and can be consistently estimated non-parametrically using the so-called Aalen-Johansen estimator (Aalen and Johansen 1978). However, in many cases, the Markov models do not fit satisfactorily, and fortunately, alternative estimators exist. Estimators which do not rely on the Markov condition were introduced for the first time by Meira-Machado, de Uña-Álvarez, and Cadarso-Suárez (2006). The proposed estimators were defined using multivariate Kaplan-Meier integrals employing time multivariate techniques. These authors showed that the proposed estimators may behave more efficiently than the Aalen-Johansen estimators in case of strong violation of the Markov condition. This work has been recently revisited by de Uña-Álvarez and Meira-Machado (2015) who proposed alternative estimators based on subsampling which are consistent regardless the Markov condition. The idea behind their estimators is to use specific subsamples or portions of the data at hand for which the ordinary Kaplan and Meier (1958) survival function leads to a consistent estimator of the target. The same idea of subsampling, combined with the Aalen-Johansen estimate of the state occupation probabilities derived from the same subsets, was later used by Putter and Spitoni (2018) to introduce new estimators which were termed as landmark Aalen-Johansen.
The idea behind the landmark estimators is based on the computation of the so-called occupation probabilities in a subset of individuals that happen to be in a given state at a particular time. Because of this, in some cases the computation of the transition probabilities are based on reduced data and usually in the presence of heavily censored data leading to estimators with higher variability. To avoid this problem, Meira-Machado (2016) propose two approaches that can be used to reduce the variability of the landmark estimator (de Uña-Álvarez and Meira-Machado 2015). One approach is based on spline smoothing while the other is based on a preliminary parametric estimation (presmoothing) of the probability of censoring. The main idea of presmoothing is that each censoring indicator is replaced by a smooth fit of a binary regression of the indicator on observables. The use of presmoothed estimators is a good alternative in these situations, since they give mass to all the event times, including the censored observations. Successful applications of presmoothed estimators include estimation of the survival function (Dikta 1998; Meira-Machado, Sestelo, and Gonçalves 2016), nonparametric curve estimation (Cao and Jacome 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, de Uña-Álvarez, and Meira-Machado 2011), and the estimation of the transition probabilities (Amorim, de Uña-Álvarez, and Meira-Machado 2011). All these references concluded that the presmoothed estimators have improved variance when compared to purely nonparametric estimators.

In this work we review some recent developments in this area, focusing on the presmoothed estimation of the transition probabilities. Specifically, we focus on the choice of the preliminary smoothing function which may be based on a certain parametric family such as the logistic, probit etc., or on a nonparametric regression smoother such as the Nadaraya-Watson kernel estimator (Wand and Jones 1995).

The organization of the paper is as follows. In Sec. 2 we introduce the quantities to estimate as well as the corresponding estimators. The performance of four sets of estimators is investigated through simulations in Sec. 3, while in Sec. 4 the methods are compared through the analysis of medical data from a controlled clinical trial in liver cirrhosis. In Sec. 5 we give a brief overview of the R package developed by the authors. Main conclusions are reported in Sec. 6.

2. Presmoothed estimators of the transition probabilities

2.1. Notation

A multi-state model is a model for a time continuous stochastic process \((Y(t), t \in [0, \infty))\) which at any time occupies one of a few possible states. In this paper we consider the progressive illness-death model depicted in Figure 1, i.e., we consider the state space \(\{0, 1, 2\}\), right-continuous sample paths and we assume that all subjects are in State 0 at time \(t = 0\), i.e., \(P(Y(0) = 0) = 1\). This model is encountered in many medical studies for describing the progression of patients undergoing a given illness, particularly in cancer studies.

We may define the sojourn time in the initial state, \(Z = \inf\{t : Y(t) \neq 0\}\) and the total time of the process \(T = \inf\{t : Y(t) = 2\}\). Note that \((Z, T)\) falls on a line with a
strictly positive probability, since \( T = Z \) for those individuals undergoing a direct transition from State 0 to the absorbing State 2. On the other hand, \( Z < T \) indicates that the individual visits the intermediate State 1 at some time. In practice, several issues influence the observation of these two random variables. The most common issue is right-censoring, which happens when a subject leaves the study before an event occurs, or when the study ends before the event has occurred. Under right-censoring, only the censored versions of \( Z \) and \( T \), along with their corresponding censoring indicators, are available. This censoring is modeled by considering a censoring variable \( C \), which we assume to be independent of the process \((Z, T)\). Define \( \tilde{Z} = \min(Z, C) \) and \( \tilde{T} = \min(T, C) \) for the censored versions of \( Z \) and \( T \) and introduce \( \Delta_1 = I(Z \leq C) \) and \( \Delta = I(T \leq C) \) for the respective censoring indicators of \( Z \) and \( T \). The variables \( \tilde{Z} \) and \( \tilde{T}_{12} = \tilde{T} - \tilde{Z} \) are the observed sojourn times in states 0 and 1, respectively. Finally, the available data are \((\tilde{Z}_i, \tilde{T}_i, \Delta_{1i}, \Delta_i)\), \(1 \leq i \leq n\), i.i.d. copies of \((\tilde{Z}, \tilde{T}, \Delta_1, \Delta)\).

### 2.2. Estimation of the transition probabilities

For two states \( h, j \) and two time points \( s < t \), introduce the so-called transition probabilities \( p_{hj}(s, t) = P(Y(t) = j | Y(s) = h, \mathcal{H}_s) \), where \( \mathcal{H}_s \) denotes the history of the multi-state process up to \( s \). The history of the process will include information such as the states previously visited, transition times, etc. The process is Markovian, if for any \( s, t \) with \( 0 \leq s < t \), and \( h, j, y \in S \), the state space of the process, we have \( P(Y(t) = j | Y(s) = h, Y(u) = y) = P(Y(t) = j | Y(s) = h), 0 \leq u < s \), and thus, the future of the process after time \( s \) depends only on the state occupied at time \( s \), not on the arrival time to that state or on the states previously visited. Similarly, in Markov models, the transition intensities depend on the individual’s past history only through the current state.

Traditionally, the Markov condition is verified by modeling particular transition intensities on aspects of the history of the process using a proportional hazard model (Kay 1986). In the progressive illness-death model the Markov condition is only relevant for the transition from the disease state (State 1) to death (State 2). Under this model, we can examine whether the time spent in the initial state (State 0) is important on the transition from the disease state to death. For doing that, let \( \lambda_{12}(t) \) denote the hazard function of \( T \) for those individuals going from State 1 to State 2. Fitting a Cox model \( \lambda_{12}(t|Z) = \lambda_{12,0}(t) \exp(\beta Z) \), where \( \lambda_{12,0} \) is the baseline hazard and \( \beta \) a regression parameter, we now need to test the null hypothesis, \( H_0 : \beta = 0 \), against the general alternative, \( H_1 : \beta \neq 0 \). This would assess if the transition rate from the disease state into death is unaffected by the time spent in the disease state.

In the illness-death model, the five transition probabilities can be expressed as depending on the joint distribution of \((Z, T)\) as follows

\[
\begin{align*}
    p_{00}(s, t) &= P(Z > t | Z > s), \\
    p_{01}(s, t) &= P(Z \leq t, T > t | Z > s), \\
    p_{02}(s, t) &= P(T \leq t | Z > s), \\
    p_{11}(s, t) &= P(Z \leq t, T > t | Z \leq s, T > s), \\
    p_{12}(s, t) &= P(T \leq t | Z \leq s, T > s)
\end{align*}
\]

Recently, de Uña-Álvarez and Meira-Machado (2015) have used the idea of subsampling to introduce (landmark) estimators of the transition probabilities which do not
rely on the Markov property. The idea of the new methods is to use a procedure based on (differences between) Kaplan-Meier estimators derived from a subset of the data consisting of all subjects observed to be in a given state at a given time. Following equations above, given the time point \( s \), to estimate \( p_{0j}(s,t) \) for \( j = 0, 1, 2 \) the landmark analysis is restricted to the individuals observed in State 0 at time \( s \). For the subpopulation \( Z > s \), the censoring time \( C \) is still independent of the pair \( (Z, T) \) and, therefore, Kaplan-Meier-based estimation will be consistent. Similarly, to estimate \( p_{1j}(s,t) \), \( j = 1, 2 \), the landmark analysis proceeds from the sample restricted to the individuals observed in State 1 at time \( s \). Then, we may formally introduce the landmark estimators as follow

\[
\hat{p}^{LM}_{00}(s,t) = \hat{S}^{KM(s)}_0(t) \hat{p}_{11}(s,t) = \hat{S}^{KM(s)}_1(t) \\
\hat{p}^{LM}_{01}(s,t) = \hat{S}^{KM(s)}_0(t) - \hat{S}^{KM(s)}_0(t) \hat{p}^{LM}_{12}(s,t) = 1 - \hat{S}^{KM(s)}_1(t) \\
\hat{p}^{LM}_{02}(s,t) = 1 - \hat{S}^{KM(s)}_1(t)
\]

where \( \hat{S}^{KM(s)}_0 \) and \( \hat{S}^{KM(s)}_1 \) are the Kaplan-Meier estimators for the distributions of \( Z \) and \( T \), respectively, but computed from the subsample \( S_1 = \{ i : \tilde{Z}_i > s \} \); whereas \( \hat{S}^{KM(s)}_2 \) is the Kaplan-Meier estimator of the distribution of \( T \) but computed from the subsample \( S_2 = \{ i : \tilde{Z}_i \leq s < T_i \} \).

As can be seen in Andersen et al. (1993), the Kaplan-Meier estimator play an important role for the landmark estimators proposed by de Uña-Álvarez and Meira-Machado (2015). Under our notation, the Kaplan-Meier product-limit estimator of the survival function \( S(t) = P(T > t) \) can be expressed using Kaplan-Meier weights as follows

\[
\hat{S}^{KM}(t) = 1 - \sum_{i=1}^{n} W_i I(\tilde{T}_{(i)} \leq t)
\]

where

\[
W_i = \frac{\Delta_{[i]}}{n - i + 1} \prod_{j=1}^{i-1} \left[ 1 - \frac{\Delta_{[j]}}{n - j + 1} \right]
\]

is the Kaplan-Meier weight attached to \( \tilde{T}_{(i)} \).

Similarly, one could introduce the Kaplan-Meier formula based on the \( (\tilde{Z}_i, \Delta_{[i]}) \)'s for the distribution of \( Z \).

The Kaplan-Meier estimator is a step function with jumps located only at the uncensored observations. In heavily censored data sets the accuracy of their estimation might not be acceptable in particular at the right tail of the distribution. This problem becomes worse when this estimator is applied to subsamples of the complete data. Because of this, the landmark estimators introduced in Andersen et al. (1993) may result in wiggly estimators with fewer jump points and high variability. These problems will be more prominent in some transition probabilities. To handle these difficulties, the use of presmoothed Kaplan-Meier estimators in Andersen et al. (1993) are a good alternative since these estimators will give mass to all the event times, including the censored observations. The presmoothed Kaplan-Meier estimators are obtained by replacing the censoring indicator variables in the expression of the Kaplan-Meier weights with some smooth fit before the Kaplan-Meier formula is applied. The presmoothed Kaplan-Meier estimator of the survival function of \( T \) is given by
\[ \hat{S}_{\text{PrKM}}(t) = 1 - \sum_{i=1}^{n} \hat{W}_i I(\hat{T}(i) \leq t), \]

where

\[ \hat{W}_i = \frac{p_n(\hat{T}(i))}{n - i + 1} \prod_{j=1}^{i-1} \left[ 1 - \frac{p_n(\hat{T}(j))}{n - j + 1} \right], \]

and \( p_n(t) \) stands for an estimator of the binary regression function \( p(t) = P(\Delta = 1 | T = t) \), i.e., the conditional probability that the observation at time \( t \) is not censored.

We may now formally introduce the presmoothed landmark estimators as follows

\[ \begin{align*}
\hat{p}_{00} \text{PrLM}(s, t) &= \hat{S}_{0}^{\text{PrKM}(s)}(t), \\
\hat{p}_{11} \text{PrLM}(s, t) &= \hat{S}^{\text{PrKM}[i]}(t) \\
\hat{p}_{01} \text{PrLM}(s, t) &= \hat{S}^{\text{PrKM}(s)}(t) - \hat{S}_{0}^{\text{PrKM}(s)}(t), \\
\hat{p}_{12} \text{PrLM}(s, t) &= 1 - \hat{S}^{\text{PrKM}[i]}(t)
\end{align*} \]

where \( \hat{S}_{0}^{\text{PrKM}(s)} \) and \( \hat{S}^{\text{PrKM}(s)} \) are the presmoothed Kaplan-Meier estimators for the distributions of \( Z \) and \( T \), respectively, but computed from the respective subsamples.

One useful parametric candidate for the binary regression function \( p(t) \) belongs to a parametric family of binary regression curves, such as logit or probit. When the parametric model specified for \( p(t) \) is correct the corresponding semiparametric presmoothed estimator is at least as efficient as the original nonparametric Kaplan-Meier estimator. Importantly, the validity of a given model for the presmoothing function can be checked graphically or formally, by applying a goodness-of-fit test such as the test proposed by Hosmer, Lemeshow, and Sturdivant (2013) for the logistic model or the Kolmogorov-Smirnov type version of the model-based bootstrap approach described in Dikta, Kvesic, and Schmidt (2006). This implies that the risk of a misspecified model can be controlled in practice. An alternative and more flexible approach is to model the binary regression function through an additive regression model.

Nonparametric presmoothing (Cao et al. 2005) is useful when there is a clear risk of a misspecification of the parametric model. In this case, one may use the Nadaraya-Watson kernel estimator for \( p(\cdot) \) based on the binary responses \( \Delta_i \) with covariates \( T_i \). The idea is to estimate \( p(\hat{T}_i) \) by \( p_n(\hat{T}_i) \) where

\[ p_n(t) = \frac{\sum_{i=1}^{n} K((t - T_i)/a_n)\Delta_i}{\sum_{i=1}^{n} K((t - T_i)/a_n)} \]

and \( K \) is a known probability density function (the kernel function) and \( a_n \) is a sequence of bandwidths. As a drawback, nonparametric regression requires the specification of a bandwidth for the computation of the smooth fit \( p_n(t) \). As \( a_n \) decreases the roughness of the resulting estimator will increase. When \( a_n \to 0 \) then the presmoothed estimator coincide in the limit with the classical estimators. On the other hand, as the bandwidth \( a_n \) increase oversmoothed estimates will be obtained removing important features of the underlying structure of the survival function to be estimated.

Different presmoothed estimators are obtained depending on the choice of the preliminary smoothing function. Simulation studies reported in the next section show that in both cases the presmoothed estimators may be much more efficient than the
Table 1. Bias and standard deviation (SD) for estimators of $p_{ij}(s,t)$. The relative MSEs are also given. Scenario 1: illness-death model with correlated exponential gap times.

| (s,t) | $p_{ij}^{LM}(s,t)$ bias (SD) | $p_{ij}^{PLM}(s,t)$ bias (SD) | $p_{ij}^{NP}(s,t)$ bias (SD) | $p_{ij}^{GAM}(s,t)$ bias (SD) | Relative MSE |
|-------|-----------------------------|-------------------------------|-------------------------------|-------------------------------|---------------|
|       | n                           | C                             | LM/PLM                        | LM/NP                         | LM/GAM        |
|       | 50                          | U[0.4]                        | 0.0001 (0.0585)               | -0.0074 (0.0521)              | -0.0079 (0.0499) | -0.0039 (0.0543) | 1.2333 | 1.3405 | 1.1516 |
|       | U[0.3]                      | 0.0024 (0.0593)               | -0.0063 (0.0515)              | -0.0075 (0.0480)              | 0.0007 (0.0531)  | 1.3039 | 1.4922 | 1.2464 |
|       | 100                         | U[0.4]                        | -0.0015 (0.0411)              | -0.0089 (0.0367)              | -0.0092 (0.0355)              | -0.0045 (0.0377) | 1.1896 | 1.2575 | 1.1770 |
|       | U[0.3]                      | 0.0010 (0.0424)               | -0.0074 (0.0360)              | -0.0084 (0.0350)              | -0.0030 (0.0387)              | 1.3285 | 1.3870 | 1.1916 |
|       | 250                         | U[0.4]                        | 0.0011 (0.0247)               | -0.0063 (0.0219)              | -0.0065 (0.0215)              | -0.0019(0.0225) | 1.1850 | 1.2175 | 1.2032 |
|       | U[0.3]                      | 0.0007 (0.0265)               | -0.0068 (0.0226)              | -0.0081 (0.0226)              | -0.0020 (0.0238)              | 1.2597 | 1.2252 | 1.2329 |
|       | (s,t)=(0.2231,0.9163)       | n                             | C                             | LM/PLM                        | LM/NP                         | LM/GAM        |
|       | 50                          | U[0.4]                        | 0.0012 (0.0746)               | -0.0027 (0.0660)              | 0.0003 (0.0644)               | <0.0001 (0.0682) | 1.2777 | 1.3446 | 1.1974 |
|       | U[0.3]                      | 0.0031 (0.0751)               | -0.0090 (0.0655)              | 0.0009(0.0641)                | 0.0012(0.0693)                | 1.3189 | 1.3760 | 1.1778 |
|       | 100                         | U[0.4]                        | -0.0009 (0.0526)              | -0.0058 (0.0465)              | -0.0022 (0.0463)              | -0.0019 (0.0477) | 1.2582 | 1.2848 | 1.2130 |
|       | U[0.3]                      | 0.0018 (0.0539)               | -0.0017 (0.0460)              | 0.0027 (0.0466)               | 0.0008 (0.0489)               | 1.3722 | 1.3777 | 1.2196 |
|       | 250                         | U[0.4]                        | 0.0016 (0.0324)               | -0.0029 (0.0288)              | 0.0012 (0.0293)               | 0.0011 (0.0296) | 1.2532 | 1.2281 | 1.1984 |
|       | U[0.3]                      | -0.0002 (0.0336)              | -0.0033 (0.0287)              | 0.0011(0.0300)                | -0.0002 (0.0301)              | 1.3540 | 1.2482 | 1.2479 |
|       | (s,t)=(0.5108,1.6094)       | n                             | C                             | LM/PLM                        | LM/NP                         | LM/GAM        |
|       | 50                          | U[0.4]                        | 0.0007 (0.0988)               | 0.0018 (0.0870)               | 0.0064 (0.0880)               | 0.0008 (0.0919) | 1.2866 | 1.2539 | 1.1514 |
|       | U[0.3]                      | 0.0018 (0.1112)               | 0.0079 (0.0930)               | 0.0030 (0.0959)               | 0.0013 (0.0987)               | 1.3920 | 1.3214 | 1.2682 |
|       | 100                         | U[0.4]                        | -0.0003 (0.0709)              | -0.0029 (0.0620)              | 0.0042 (0.0634)               | -0.0030 (0.0649) | 1.3101 | 1.2493 | 1.1931 |
|       | U[0.3]                      | 0.0040 (0.0756)               | 0.0070 (0.0640)               | 0.0017 (0.0665)               | 0.0046 (0.0680)               | 1.3827 | 1.2369 | 1.2335 |
|       | 250                         | U[0.4]                        | <0.0001 (0.0458)              | <0.0001 (0.0398)              | 0.0070 (0.0410)               | 0.0005 (0.0416) | 1.3271 | 1.2141 | 1.2133 |
|       | U[0.3]                      | -0.0002 (0.0485)              | 0.0027 (0.0394)               | 0.0109 (0.0424)               | 0.0005 (0.0425)               | 1.5079 | 1.2224 | 1.2975 |
|       | (s,t)=(0.9163,1.6094)       | n                             | C                             | LM/PLM                        | LM/NP                         | LM/GAM        |
|       | 50                          | U[0.4]                        | 0.0041 (0.1332)               | -0.0013 (0.1201)              | -0.0016 (0.1178)              | -0.0006 (0.1199) | 1.2304 | 1.2801 | 1.2347 |
|       | U[0.3]                      | -0.0016 (0.1422)              | -0.0042 (0.1279)              | -0.0095 (0.1273)              | -0.0182 (0.1254)              | 1.2345 | 1.2428 | 1.2643 |
|       | 100                         | U[0.4]                        | -0.0043 (0.0848)              | -0.011 (0.0760)               | -0.0082 (0.0766)              | -0.0075 (0.0783) | 1.2226 | 1.2152 | 1.1629 |
|       | U[0.3]                      | -0.0020 (0.1014)              | -0.0027 (0.0872)              | -0.0002 (0.0893)              | -0.0069 (0.0882)              | 1.3507 | 1.2906 | 1.3134 |
|       | 250                         | U[0.4]                        | -0.0014 (0.0550)              | -0.0070 (0.0493)              | -0.0046 (0.0487)              | -0.0032 (0.0503) | 1.2330 | 1.2666 | 1.1933 |
|       | U[0.3]                      | -0.0025 (0.0586)              | -0.0044 (0.0507)              | -0.0011 (0.0530)              | -0.0030 (0.0532)              | 1.3314 | 1.2272 | 1.2132 |
| (s,t) = (0.2231, 0.5108) | n | C | U[0.4] | U[0.3] | U[0.2] |
|---|---|---|---|---|---|
| 50 | -0.0192 (0.2470) | -0.0218 (0.2452) | -0.0026 (0.2189) | 0.0219 (0.1982) | 1.0124 | 1.2812 | 1.5505 |
| 100 | -0.0087 (0.2427) | -0.0163 (0.2404) | 0.0217 (0.2124) | 0.0559 (0.1830) | 1.0159 | 1.2941 | 1.6226 |
| 250 | -0.0063 (0.1666) | -0.0112 (0.1655) | -0.0066 (0.1657) | -0.0006 (0.1592) | 1.0096 | 1.0113 | 1.0968 |
| (s,t) = (0.2231, 0.9163) | n | C | U[0.4] | U[0.3] | U[0.2] |
| 50 | -0.0026 (0.2291) | -0.0082 (0.2250) | 0.0124 (0.1918) | 0.0401 (0.1590) | 1.0357 | 1.4215 | 1.9630 |
| 100 | 0.0035 (0.2389) | -0.0137 (0.2458) | 0.0475 (0.1903) | 0.0579 (0.1558) | 0.9416 | 1.4848 | 2.0796 |
| 250 | -0.0005 (0.1624) | -0.0039 (0.1541) | -0.0016 (0.1567) | 0.0084 (0.1437) | 1.1097 | 1.0738 | 1.2728 |
| (s,t) = (0.5108, 1.6094) | n | C | U[0.4] | U[0.3] | U[0.2] |
| 50 | -0.0064 (0.1801) | -0.0120 (0.1686) | 0.0020 (0.1517) | 0.0212 (0.1378) | 1.1366 | 1.4113 | 1.6733 |
| 100 | -0.0182 (0.1910) | -0.0412 (0.1997) | 0.0028 (0.1574) | 0.0442 (0.1290) | 0.8852 | 1.4862 | 1.9947 |
| 250 | 0.0013 (0.1253) | -0.0032 (0.1179) | 0.0012 (0.1188) | 0.0004 (0.1174) | 1.1280 | 1.1115 | 1.1394 |
| (s,t) = (0.9163, 1.6094) | n | C | U[0.4] | U[0.3] | U[0.2] |
| 50 | 0.0011 (0.1953) | -0.0020 (0.1908) | 0.0076 (0.1784) | 0.0326 (0.1529) | 1.0485 | 1.1968 | 1.5640 |
| 100 | -0.0002 (0.2146) | -0.0135 (0.2114) | 0.0144 (0.1921) | 0.0103 (0.1438) | 1.0263 | 1.2414 | 1.4827 |
| 250 | 0.0007 (0.1412) | -0.0001 (0.1360) | -0.0057 (0.1409) | 0.0028 (0.1366) | 1.0782 | 1.0026 | 1.0675 |

**Note:** The table contains values that seem to be related to statistical analysis, possibly representing means or variances with standard errors or confidence intervals. The bounds seem to be 50, 100, and 250, with the columns labeled as U[0.4], U[0.3], U[0.2], which could indicate different threshold values or levels of significance.
completely nonparametric estimator, since they often have less variance while providing smoother curves with the expected behavior.

3. Simulation studies

In this section we report results from two simulation studies, where the aim is to compare the finite sample performance of our presmoothed estimators of the transition probabilities. To be specific, we compare the original landmark estimator of de Uña-Alvarez and Meira-Machado (2015) (labeled as LM) with the semiparametric presmoothed estimator which is obtained through the use of a logistic regression model for the presmoothing function (labeled as PLM) and the presmoothed estimator based on a nonparametric kernel regression model (labeled as NP). For completeness purposes we also included the presmoothed estimator which is obtained through the use of a generalized additive logistic regression model (labeled as GAM).

We first simulated data from a scenario used by Amorim, de Uña-Alvarez, and Meira-Machado (2011), which these authors found to be challenging both in terms of bias and variance. To be specific, we separately consider the subjects passing through State 1 at some time and those who directly go to the State 2. For the first subgroup of individuals we generated replicates of \( F_{12}(x,y) = F_1(x)F_2(y) \left[ 1 + \left\{ 1 - F_1(x) \right\} \left\{ 1 - F_2(y) \right\} \right] \) with exponential marginal distribution functions with rate parameter 1. For the second subgroup of individuals \( (Z = T) \), the value of \( Z \) is simulated according to an exponential with rate parameter 1. Random censoring was simulated from uniform distributions \( U[0, \tau_G] \) for \( \tau_G \) equal to 3 and 4. The model with \( \tau_G = 4 \) results in 24% of censoring on the first gap time \( Z \), and in 47% of censoring on the second gap time \( T - Z \). The model with \( \tau_G = 3 \) increases these censoring levels to 32% and about 57%, respectively. Since we are assuming correlated times for \( Z \) and \( T - Z \), the simulated model does not satisfies the Markov property. Details of the simulation procedure can be found in Amorim, de Uña-Alvarez, and Meira-Machado (2011).

To summarize the results we fixed the values of \( (s, t) \) for four different points, corresponding to combinations of the percentiles 20%, 40%, 60% and 80% of the exponential marginal distribution functions with rate parameter 1. In each simulation, 1000 samples were generated with sample sizes 50, 100 and 250. As a measure of efficiency, we took the Mean Squared Error (MSE) but we also computed the standard deviations (SD) and the Bias for each point \( (s, t) \).

The performance of the empirical transition probabilities \( \hat{p}_{ij}^{LM}(s,t), \hat{p}_{ij}^{PLM}(s,t), \hat{p}_{ij}^{GAM}(s,t) \) and \( \hat{p}_{ij}^{NP}(s,t) \) in the simulations are summarized in Table 1. The semiparametric presmoothing (labeled as PLM) requires to estimate the binary regression functions, such as \( p_{0n}(t) = P(\Delta_1 = 1 | \tilde{Z} = t) \) and \( p_{1n}(t) = P(\Delta = 1 | \tilde{T} = t) \), to presmooth the Kaplan-Meier estimators \( \hat{S}_0(t) \) and \( \hat{S}(t) \), respectively. After some algebra, it is seen that the function \( p_{0n}(t) \) is written as \( (\tau_G - t)/(\tau_G - t + 1) \). The binary function \( p_{1n}(t) \), for those individuals that observe a transition to State 1, is given by \( \tau_G/(1 + \eta(t)) \), where \( \eta(t) = \lambda_G(t)/\lambda_T(t) \), and where \( \tau_G(t) = 1/(\tau_G - t) \) is the hazard rate of the censoring variable and

\[
\lambda_T(t) = \left( -e^{-t}(4 - 2t) + e^{-2t}(8 + 4t) \right) / \left( e^{-t}(2t - 2) + e^{-2t}(3 + 2t) \right)
\]
is the hazard rate of $T$ under restriction $\tilde{Z} < \tilde{T}$. Plots for these functions shown in Figure 2 reveal that they are monotonous and so they can be adequately estimated by logistic regression.

In our simulations we have used the logistic regression models $p_n(t; \beta) = 1/(1 + \exp(\beta_0 + \beta_1 t))$. The $\beta$ parameter in model $p_{0n}(\cdot; \beta)$ was estimated by maximizing of the conditional likelihood of the $\Delta_i$’s given $\tilde{Z}$; and via maximization of the conditional likelihood of the $\Delta$’s given $\tilde{T}$, in model $p_{1n}(\cdot; \beta)$. The presmoothing labeled as GAM was implemented by fitting a generalized additive logistic model through the mgcv R package (Wood 2019). To compute the transition probabilities with a nonparametric presmoothing we have used the plug-in bandwidth selector of Cao et al. (2005) and biweight kernels. As would be expected, results reported in this table reveal that the estimation of the transition probabilities is performed with less accuracy as $s$ and $t$ grow. This behavior was expected since the performance of all methods in lifetime data is usually poorer at the right tail where the censoring effects are stronger. At these points the SD is in most cases larger. The SD decreases with an increase in the sample size and with a decrease of the censoring percentage. All methods obtain in all settings a low bias. It can also be seen that the SD clearly dominates the performance of the proposed estimators in most the cases revealing a clear advantage of the presmoothed estimators when compared with the unsmoothed landmark estimators (labeled as LM). This can be observed by the relative efficiency between the unsmoothed landmark estimator and the three presmoothed estimators that was measured by the ratio between their corresponding MSEs. In almost all cases, the use of presmoothing leads to estimators with less SD and less MSE.

Because of space limitation, Table 1 does not report the results for transition probability $\hat{p}_{00}(s,t)$. However, the behavior of the estimators for this transition can be seen in Figures 3 and 4 through the boxplots of the mean squared errors based on the 1000 Monte Carlo replicates for the four estimators, with different sample sizes and two censoring levels. For completeness purposes we decided to show the plots for three transitions and different fixed values of $(s, t)$. The boxplots shown in these figures reveal some results which are in agreement with our findings reported in Table 1. These plots confirm the less variability of the presmoothed estimators. Results shown in Table 1
Figure 3. Boxplots of the $M = 1000$ estimates of the transition probabilities. Horizontal solid red line correspond to the true value of the transition probability. Censoring times uniformly distributed between 0 and 3.
Figure 4. Boxplots of the $M = 1000$ estimates of the transition probabilities. Horizontal solid red line correspond to the true value of the transition probability. Censoring times uniformly distributed between 0 and 4.
and Figures 3 and 4 suggest that the use of presmoothing leads to better results for all transition probabilities while neither one of the three presmoothed estimators (PLM, GAM and NP) seems to be uniformly the best.

While reducing the variance, presmoothing may introduce some bias in estimation. Simulation results reported in Table 1 serve to illustrate this issue too. When the parametric model specified for the presmoothing function is incorrect, the corresponding semiparametric estimator may lose efficiency, providing estimators with a large bias. However, the validity of a given parametric model for presmoothing can be checked by applying a goodness-of-fit test such as the test proposed by Hosmer, Lemeshow, and Sturdivant (2013). We have conducted simulation studies to evaluate the capability of the Hosmer and Lemeshow goodness-of-fit test to detect deviations from the logistic regression model. Results (not reported here) suggest small deviations in this scenario. In most cases the test was not able to reject the logistic model. Though the simulated scenario seem to be favorable to the estimator based on a parametric preliminary smoothing, the estimator based on nonparametric presmoothing is competitive, attaining better results (with less variance and less mean squared errors) in a large number of points.

To assess the effect of a misspecification of the logistic regression model in the estimation of the transition probabilities through the use of parametric presmoothing methods, a second simulation study was performed. In this scenario, data were generated according to the progressive three-state model. This model can be viewed as a particular case of the illness-death model where no transitions are observed on disease-free mortality transition. The vector of gap times \((Z, T - Z)\) is simulated as follows. The first gap time \(Z\) is simulated according to a mixture of two lognormal distributions, \(\text{LogN}(0, 3)\) and \(\text{LogN}(3, 0.2)\), with equal probability. Given \(Z = x\), the second gap time, \(T_{12} = T - Z\), is drawn from a Lognormal distribution with mean \(\log(x)/2\) and standard deviation equal to 0.2. This scenario does not follow the Markov assumption since the hazard for the second gap time depend on the time to progression to the intermediate state. The censoring time \(C\) was independently generated following a uniform distribution \(U[10, 30]\). Note that under this scenario, censored observations are expected to be concentrated near the center of the distribution and therefore revealing a misspecification of the fitted logistic regression model.

Results shown in Table 2 reveal the impact of a misspecification of the fitted logistic regression model in the estimation of the transition probabilities, leading to estimators with increased bias, in some points, and higher values for the MSE. In such cases the use of an estimator based on nonparametric presmoothing is preferable.

4. Example of application

For illustration purposes, we apply the proposed methods of Sec. 2 to data based on a case study presented by the PROVA study group which aimed to evaluate the effect of propranolol and sclerotherapy on risk of first variceal bleeding and survival in patients with liver cirrhosis. The trial included 286 patients in whom cirrhosis has been diagnosed in the past and where endoscopy had shown esophageal varices but who had not yet experienced a transfusion-requiring bleeding from the varices (PROVA Study Group 1991). From the 286 patients
enrolled in the trial, 50 had a bleeding episode and among these 29 died. Forty six patients died without experienced variceal bleeding. The rest of the patients (190) remained alive and without variceal bleeding up to the end of the follow-up. The patients were observed for up to 42 months, with an average of 15 months. For each patient, the two event times (time to variceal bleeding and time to death) and the corresponding indicator statuses are recorded. The process can be modeled using a progressive illness-death model with transient states.

The dataset for the PROVA study group is of small size with a high censoring percentage, near to 74%. The time alive and without variceal bleeding, $Z$, is more often censored in the time interval between 1 and 4 years. We have only 11 individuals that experienced variceal bleeding one year after enrollment in the study. A few (de Uña-

### Table 2. Bias and standard deviation (SD) for estimators of $p_{ij}(s,t)$. The relative MSEs are also given.

| $(s,t)$ | $\hat{p}_{ij}^{LM}(s,t)$ bias (SD) | $\hat{p}_{ij}^{PLM}(s,t)$ bias (SD) | $\hat{p}_{ij}^{NP}(s,t)$ bias (SD) | $\hat{p}_{ij}^{GAM}(s,t)$ bias (SD) | Relative MSE |
|--------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------|
| (8, 10) | $n$ | $100$ | $250$ | $500$ | $n$ | $100$ | $250$ | $500$ | $n$ | $100$ | $250$ | $500$ |
| 100 | $-0.0004$ | $0.0004$ | $-0.0002$ | $0.0002$ | $0.0178$ | $0.0108$ | $0.0178$ | $0.0178$ | $0.0131$ | $1.1609$ | $1.0168$ | $1.0599$ |
| 250 | $-0.0140$ | $0.0114$ | $-0.0147$ | $0.0161$ | $0.0178$ | $0.0178$ | $0.0178$ | $0.0178$ | $0.0131$ | $1.1609$ | $1.0168$ | $1.0599$ |
| 500 | $-0.0079$ | $0.0091$ | $-0.0025$ | $0.0096$ | $0.0210$ | $0.0108$ | $0.0210$ | $0.0210$ | $0.0073$ | $1.2339$ | $1.2644$ | $0.5793$ | $1.0502$ | $1.1486$ | $0.9700$ | $0.9913$ | $1.0240$ |
| 100 | $-0.0004$ | $0.0072$ | $-0.0082$ | $0.0021$ | $0.0273$ | $0.0273$ | $0.0273$ | $0.0273$ | $0.0703$ | $1.4298$ | $1.5724$ | $1.2729$ |
| 250 | $-0.0013$ | $0.0014$ | $-0.0069$ | $0.0021$ | $0.0273$ | $0.0273$ | $0.0273$ | $0.0273$ | $0.0703$ | $1.4298$ | $1.5724$ | $1.2729$ |
| 500 | $-0.0010$ | $0.0013$ | $-0.0042$ | $0.0010$ | $0.0273$ | $0.0273$ | $0.0273$ | $0.0273$ | $0.0703$ | $1.4298$ | $1.5724$ | $1.2729$ |

The relative MSEs are also given.

The process can be modeled using a progressive illness-death model with transient states ‘alive and without variceal bleeding’, ‘alive with variceal bleeding’, and an absorbing state ‘dead’.

The dataset for the PROVA study group is of small size with a high censoring percentage, near to 74%. The time alive and without variceal bleeding, $Z$, is more often censored in the time interval between 1 and 4 years. We have only 11 individuals that experienced variceal bleeding one year after enrollment in the study. A few (de Uña-
Álvarez and Rodríguez-Campos (2004) complete (i.e., uncensored) observations for patients with a survival time greater than one year that died without experiencing variceal bleeding. At 90 days (3 months) we have 260 individuals alive and without variceal bleeding and only 11 individuals alive with variceal bleeding. From those 11 that experiencing variceal bleeding only 3 died leading to 73% of censored observations. Similarly, at 1 year after enrollment, we have 21 individuals alive with variceal bleeding and a few 5 complete observations (76% censoring). This means that it is expected that any estimator for the transition probabilities will behave poorly in these cases. Presmoothing will help to improve the estimation in these situations, in particular in the right tail of the distributions of $T$ and $T-Z$ (i.e., for larger times for being alive and without variceal bleeding, and also for larger values of the time since variceal bleeding, $T-Z$).

It is of practical interest to determine whether the Markov condition holds within a particular dataset in order to determine which method is more appropriate to estimate the transition probabilities. The Markov condition may be checked by studying the effect of the time of variceal bleeding on the mortality after bleeding. Results from fitting this covariate in a Cox model reported a significant coefficient (p-value $= 0.0003$; regression coefficient: $0.0061$) for the bleeding time revealing the failure of the Markov condition for the PROVA dataset. These findings are in agreement with those obtained by Andersen, Esbjerg, and Sorensen (2000) and Andersen and Keiding (2002).

In cases like this one, in which there is strong evidence that the process does not verifies the Markov property, the landmark estimators can be preferable due to their greater accuracy. Since the construction of these estimators is based on subsamples of the complete data, this will generally lead to small sample sizes and usually to heavily censored data as shown in Figure 5. This fact is more evident for the individuals with variceal bleeding (observed in State 1) but it is also present for those in the alive and without bleeding state (State 0) for higher landmarking times. To avoid this problem, a valid approach is to consider a modification of the landmark estimator based on presmoothing. Here we present some figures to illustrate the estimators obtained by replacing the censoring indicator variables in the classical definitions by values of several regression estimators.

Figure 5. Number of individuals observed in State 0 and State 1 along time. PROVA data.
Figure 6 shows estimated curves of the transition probabilities \( \hat{p}_{ij}(s,t) \), for fixed values of \( s = 90 \) and \( s = 365 \) days along time \( t \). The estimators labeled as ‘Unsmoothed’ correspond to the original unsmoothed landmark estimator proposed by de Uña-Álvarez and Meira-Machado (2015) whereas the estimator labeled as ‘logit’ correspond to the semiparametric presmoothed estimator introduced in Sec. 2.2 with a preliminary presmoothing based on a parametric binomial logistic family. For completeness purposes we also included semiparametric presmoothed estimators based on a different binomial family such as ‘probit’ or ‘cauchit’. We also considered estimators with a preliminary presmoothing function based on an additive logistic model (labeled as ‘logit.gam’). Finally, the estimator labeled as ‘nonparametric’ correspond to the presmoothed estimators with a preliminary presmoothing based on a nonparametric regression model using the Nadaraya-Watson kernel estimator. None of these methods require the process to be Markov.

As expected, all estimators report roughly the same estimates for lower values of \( t \) whereas some deviations are observed for higher values of \( t \). It can also be seen that the original unsmoothed landmark estimator (de Uña-Álvarez and Meira-Machado 2015) reveals higher variability on the right hand side. Plots for the transition probabilities \( \hat{p}_{00}(s,t) \) (first row) and \( \hat{p}_{11}(s,t) \) (last row), \( s = 90 \) and 365, report the survival fraction along time, among the individuals ‘alive and without variceal bleeding’ and among those with a ‘bleeding episode’. Results for the estimated curves for \( \hat{p}_{00}(s,t) \), \( s = 90 \) and 365, reveal also that the choice of the family function for the regression estimator can lead to major differences in the estimates. Some differences can also be observed in the right tail of the estimated transition probabilities \( \hat{p}_{01}(s,t) \) (second row) and \( \hat{p}_{02}(s,t) \) (third row). Plots for \( \hat{p}_{02}(s,t) \), report one minus the the survival fraction along time, among the individuals alive and without variceal bleeding at time \( s \). Plots for \( \hat{p}_{01}(s,t) \), allow for an inspection along time of the probability of being ‘alive with bleeding episode’ for the individuals who are without variceal bleeding at time \( s \). Since the ‘bleeding’ state is transient, these curves are first increasing and then decreasing. Major differences are observed in the right tail when comparing the methods based on a parametric presmoothing with their counterparts (‘unsmoothed’ and ‘nonparametric’) for the estimation of the transition probabilities. This can be possibly explained by a misspecification of the parametric model. The logistic model is likely the most common statistical model that is usually taken to apply to a binary dependent variable. The validity of a logistic model for the presmoothing function can be checked by applying a goodness-of-fit test such as the test proposed by Hosmer, Lemeshow, and Sturdivant (2013). Results obtained for all landmarking times (i.e., \( s \)) are depicted in Figure 7, revealed that the test was able to reject the logistic model when used to estimate the transition probabilities \( \hat{p}_{0j}(90,t) \). Note that the choice of this parametric model is a common choice for a parametric presmoothing. However, curves depicted in Figure 6 reveal that the presmoothed landmark estimators with a preliminary presmoothing based on a nonparametric regression provides in all cases curves with the expected behavior, similar to those obtained from the nonparametric original landmark estimators (labeled as ‘unsmoothed’) but with a better description of the tail behavior. Since there is a clear risk of a misspecification of the parametric model, the use of estimators based on nonparametric presmoothing as those labeled with ‘nonparametric’ are preferable.
Figure 6. Estimated transition probabilities for $s = 90$ and $s = 365$. PROVA data.
5. Software development

To provide the biomedical researchers with a comprehensive and easy-to-use tool for obtaining estimates of the transition probabilities we developed an R package (R Core Team 2018) called presmTP (Soutinho, Meira-Machado, and Oliveira 2019). The main function presmTP() can be used to calculate the estimated transition probabilities from the unsmoothed landmark estimators but presmoothed estimates can also be obtained through the use of a parametric family of binary regression curves, such as logit, probit or cauchit. The additive logistic regression model and nonparametric regression are also alternatives implemented in function presmTP(). Examples and details on the usage of the functions that compose the package can be obtained with the corresponding help pages. The package dependencies include the survPresmooth package (López-de-Ulibarri and Jácome 2013) and the mgcv package (Wood 2019). presmTP is written entirely in the R programming language, and is available at the CRAN repository at https://cran.r-project.org/web/packages/presmTP.

6. Conclusions

There have been several recent contributions for the estimation of the transition probabilities in the context of multi-state processes that do not satisfy the Markov property. Many of these contributions were made for the illness-death model. Recently, the problem of estimating the transition probabilities in illness-death model has been reviewed, and new (landmark) estimators have been proposed which are built by considering

Figure 7. P-values for the Hosmer and Lemeshow test for logistic regression model for all possible landmarking times. PROVA data.
specific subsets of individuals. As a weakness, the proposed estimators may provide large standard errors in estimation.

In this article the relative performance of several presmoothed landmark estimators for the transition probabilities was investigated through simulations. Attained results suggests that the use of presmoothing techniques can lead to competitive estimators that may outperform the original landmark estimators providing estimators with less variability. It is worth mentioning that presmoothing can introduce some bias in estimation while reducing the variance. This bias component is larger when there is some misspecification in the chosen parametric model. The risk of introducing a large bias through a misspecified model can be controlled in practice by testing the validity of the model applying a goodness-of-fit test. Nonparametric presmoothing is useful when there is no parametric candidate for the presmoothing function.

Software in the form of an R package was developed implementing the proposed methods.

Acknowledgements

This research was financed by Portuguese Funds through FCT - “Fundação para a Ciência e a Tecnologia”, within the research grant PD/BD/142887/2018. Luís Meira-Machado acknowledges financial support from the Spanish Ministry of Economy and Competitiveness MINECO through project MTM2017-82379-R funded by (AEI/FEDER, UE) and acronym “AFTERAM”. The Authors would also like to thank P.K. Andersen for providing the data.

ORCID

Gustavo Soutinho http://orcid.org/0000-0002-0559-1327
Luís Meira-Machado http://orcid.org/0000-0002-8577-7665
Pedro Oliveira http://orcid.org/0000-0002-2470-0795

References

Aalen, O. O., and S. Johansen. 1978. An empirical transition matrix for nonhomogeneous Markov chains based on censored observations Statistics in Medicine. *Scandinavian Journal of Statistics* 5:141–50.
Amorim, A. P., J. de Uña-Álvarez, and L. Meira-Machado. 2011. Presmoothing the transition probabilities in the illness-death model. *Statistics & Probability Letters* 81 (7):797–806. doi: 10.1016/j.spl.2011.02.017.
Andersen, P. K., and N. Keiding. 2002. Multi-state models for event history analysis. *Statistical Methods in Medical Research* 11 (2):91–115. doi: 10.1191/0962280202SM276ra.
Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding. 1993. *Statistical Models based on counting processes*. New York: Springer-Verlag.
Andersen, P. K., S. Esbjerg, and T. I. A. Sorensen. 2000. Multistate models for bleeding episodes and mortality in liver cirrhosis. *Statistics in Medicine* 19 (4):587–99. doi: 10.1002/(SICI)1097-0258(20000229)19:4<587::AID-SIM358>3.0.CO;2-0.
Cao, R., and M. A. Jácome. 2004. Presmoothed kernel density estimator for censored data. *Journal of Nonparametric Statistics* 16 (1-2):289–309. doi: 10.1080/10485250310001622622.
Cao, R., I. López-de Ullíbarri, P. Janssen, and N. Veraverbeke. 2005. Presmoothed Kaplan-Meier and Nelson-Aalen estimators. *Journal of Nonparametric Statistics* 17 (1):31–56. doi: 10.1080/10485250410001713981.
de Uña-Álvarez, J., and A. P. Amorim. 2011. A semiparametric estimator of the bivariate distribution function for censored gap times. *Biometrical Journal* 53 (1):113–27. doi: 10.1002/bimj.201000063.

de Uña-Álvarez, J., and C. Rodríguez-Campos. 2004. Strong consistency of presmoothed Kaplan-Meier integrals when covariables are present. *Statistics* 38 (6):483–96. doi: 10.1080/0233188041233136065.

de Uña-Álvarez, J., and L. Meira-Machado. 2015. Nonparametric estimation of transition probabilities in the non-Markov illness-death model: A comparative study. *Biometrics* 71 (2):364–150. doi: 10.1111/biom.12288.

Dikta, G. 1998. On semiparametric random censorship models. *Journal of Statistical Planning and Inference* 66 (2):253–79. doi: 10.1016/S0378-3758(97)00091-8.

Dikta, G., M. Kvesic, and C. Schmidt. 2006. Bootstrap approximations in model checks for binary data. *Journal of the American Statistical Association* 101 (474):521–30. doi: 10.1198/016214505000001032.

Hosmer, D. W., S. Lemeshow, and R. X. Sturdivant. 2013. *Applied logistic regression*, 3rd ed. New York: John Wiley & Sons.

Hougaard, P. 2000. *Analysis of multivariate survival data*. New York: Springer-Verlag.

Jácome, M. A., and M. C. Iglesias. 2010. Presmoothed estimation of the density function with truncated and censored data. *Statistics* 44 (3):217–34. doi: 10.1080/02331880903043231.

Kaplan, E., and P. Meier. 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53 (282):457–81. doi: 10.1080/01621459.1958.10501452.

Kay, R. 1986. A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics* 42 (4):855–65. doi: 10.2307/2530699.

López-de-Ullibarri, I., and M. A. Jácome. 2013. survPresmooth: An R Package for Presmoothed Estimation in Survival Analysis. *Journal of Statistical Software* 54 (11):1–26. doi: 10.18637/jss.v054.i11.

Meira-Machado, L. 2016. Smoothed landmark estimators of the transition probabilities. SORT-Statistics and Operations Research Transactions 40:375–98.

Meira-Machado, L., J. de Uña-Álvarez, and C. Cadarso-Suárez. 2006. Nonparametric estimation of transition probabilities in a non-Markov illness-death model. *Lifetime Data Analysis* 12 (3):325–44. doi: 10.1007/s10985-006-9009-x.

Meira-Machado, L., J. de Uña-Álvarez, C. Cadarso-Suárez, and P. K. Andersen. 2009. Multi-state models for the analysis of time to event data. *Statistical Methods in Medical Research* 18 (2):195–222. doi: 10.1177/0962280208092301.

Meira-Machado, L., M. Sestelo, and A. Gonçalves. 2016. Nonparametric estimation of the survival function for ordered multivariate failure time data: A comparative study. *Biometrical Journal* 58 (3):623–34. doi: 10.1002/bimj.201500038.

PROVA Study Group. 1991. Prophylaxis of first hemorrhage from esophageal varices by sclerotherapy, propranolol or both in cirrhotic patients: A randomized multicenter trial. *Hepatology* 14:1016–24.

Putter, H., and C. Spitoni. 2018. Non-parametric estimation of transition probabilities in non-Markov multi-state models: The landmark Aalen-Johansen estimator. *Statistical Methods in Medical Research* 27 (7):2081–92. doi: 10.1177/0962280216674497.

Putter, H., M. Fiocco, and R. B. Geskus. 2007. Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine* 26 (11):2389–430. doi: 10.1002/sim.2712.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Soutinho, G., L. Meira-Machado, and P. Oliveira. presmTP: Methods for Transition Probabilities. R package version 1.1.0. 2019. https://cran.r-project.org/web/packages/presmTP.

Wand, M. P., and M. C. Jones. 1995. *Kernel smoothing*. London, UK: Chapman and Hall.

Wood, S. 2019. mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. R Package Version 1, 8–31. https://cran.r-project.org/web/packages/mgcv.