Estimation of the Transition Probabilities in Multi-state Survival Data: New Developments and Practical Recommendations

GUSTAVO SOUTINHO
EPIUnit
University of Porto
Rua das Taipas 135, 4050-600 Porto
PORTUGAL

LUÍS MEIRA-MACHADO
University of Minho
Department of Mathematics
Campus de Azurém - 4800-058 Guimarães
PORTUGAL

Abstract: Multi-state models can be successfully used for describing complicated event history data, for example, describing stages in the disease progression of a patient. In these models one important goal is the estimation of the transition probabilities since they allow for long term prediction of the process. Traditionally these quantities have been estimated by the Aalen-Johansen estimator which is consistent if the process is Markovian. Recently, estimators have been proposed that outperform the Aalen-Johansen estimators in non-Markov situations. This paper considers a new proposal for the estimation of the transition probabilities in a multi-state system that is not necessarily Markovian. The proposed product-limit nonparametric estimator is defined in the form of a counting process, counting the number of transitions between states and the risk sets for leaving each state with an inverse probability of censoring weighted form. Advantages and limitations of the different methods and some practical recommendations are presented. We also introduce a graphical local test for the Markov assumption. Several simulation studies were conducted under different data scenarios. The proposed methods are illustrated with a real data set on colon cancer.

Key–Words: Markov assumption, Multi-state models, Nonparametric estimation, Transition probabilities, Survival Analysis

1 Introduction

Multi-state models ([1], [2], [3], [4]) are models for a stochastic process, which at any time occupies one of a set of discrete states. These models provide a relevant modeling framework to deal with complex longitudinal survival data in which individuals may experience more than one single event type. In such survival studies, besides overall survival, more than one endpoint can be observed making the use of multi-state models preferable over traditional survival methods (e.g., the Cox model and the Kaplan-Meier estimator of survival). A wide range of biomedical situations have been modeled using multi-state methods, for example, HIV infection and AIDS [5], liver cirrhosis [6], breast cancer ([7]; [8]) and problems following heart transplantation [3]. The states are usually based on clinical symptoms (e.g., bleeding episodes), biological markers (CD4 T-lymphocyte cell counts, serum immunoglobulin levels), some scale of the disease (e.g., stages of cancer or VIH infection) or a non-fatal complication in the course of the illness (e.g., heart transplantation). In cancer studies, besides death other endpoints such as locoregional recurrence and distant metastasis are often observed. A change of state is called a transition, or an event. States can be transient or absorbing. An absorbing state is a process to which one will never leave once it enters.

The state structure of a multi-state model identifies the states and also the transitions allowed between states. This structure can be represented schematically through diagrams with boxes representing the states and arrows the possible transitions that can occur. The complexity of a multi-state model greatly depends on the number of states defined and also on the transitions allowed between these states. The simplest form of a multi-state model is the mortality model with states "alive" and "dead" and a single transition allowed between them. This corresponds to the usual survival analysis situation. Splitting the "alive" state from the simple mortality model for survival data into two transient states, we therefore obtain the simplest progressive three-state model. Another possible multi-state model to describe the disease progression is the illness-death model (also known as disability model; Figure 1). In the irreversible version of this model, individuals enter the study in State 1 (e.g., healthy or alive without any disease) and subsequently move either to the transient State 2 ("diseased") or go directly

to the absorbing "dead" state. Individuals in State 2 will eventually move to the absorbing state without any possibility of recovery. Many time-to-event data sets from medical studies with multiple end points can be reduced to this generic structure.

The multi-state process can be fully characterized through its transition intensities or by its transition probabilities. The transition intensities are the instantaneous hazards for movement from one state to another. These functions can used to obtain estimates of the mean sojourn time in a given state and to determine the number of individuals in different states at a certain moment. Covariates may be incorporated in models through transition intensities to explain differences among individuals in the course of the illness. The transition probabilities represent the conditional probabilities that can be used 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 by the transition probabilities. The usual non-parametric method to estimate the transition probability matrix for non-homogeneous Markov processes is the Aalen-Johansen estimator [9]. Recently, [10] propose a modification of the Aalen-Johansen estimator in the illness-death model based on presmoothing. Although both approaches may be used to consistently estimate the occupation probabilities for non-Markov processes ([11], [12], [13]), in general they provide biased estimators for the transition probabilities if the process is not Markovian [14].

Substitute estimators for the Aalen-Johansen estimator for a non-Markov illness-death process without recovery were introduced by [14]. They showed that the alternative estimators may behave more efficiently than the Aalen-Johansen when the Markov assumption is strongly violated. Recently, [15] recovered the approach by [14] and proposed a closely related non-Markov estimator too. Both proposals ([14] and [15]) have the drawback of requiring the support of the censoring distribution to contain the support of the lifetime distribution, otherwise they only report valid estimators for truncated transition probabilities. To overcome this problem this issue was recently revisited by [16] and [17] who propose alternative estimators based on subsampling (also known as landmarking) which does not require the censoring support condition.

In this paper we revisit the topic of the nonparametric estimation of the transition probabilities, by introducing competing estimators in a multi-state system that is not necessarily Markovian and that overcomes the referred assumption on the censoring support. The new set of estimators are constructed using the cumulative hazard of the total time given a first

time but where each observation has been weighted using the information of the first duration. One of the proposed estimator is equivalent to the estimators recently proposed by [16]. To evaluate the performance of all estimators, several simulation studies were conducted under different data scenarios. Based on these results recommendations are given as to which estimators to use. We also revisit the problem of checking the Markov assumption. However, unlike the usual tests, we are interested in a more reliable test which is used for a particular transition probability. To this end, we introduce a graphical test for the Markov assumption based on the discrepancies between the new Markov-free estimators and the so-called Aalen-Johansen estimator (Markovian).
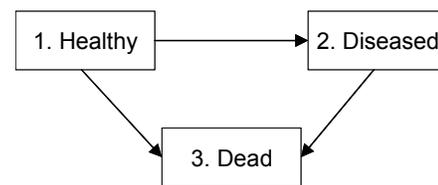

Figure 1: Illness-death model.

The organization of the paper is as follows. The estimators are introduced in Section 2. The finite sample performance of the proposed estimators for the transition probabilities is investigated via simulations in Section 3. In Section 4 we analyze data from a colon cancer study with the proposed methods. Main conclusions and specific recommendations are given in Section 5.

2 Nonparametric Estimators

2.1 Notation

A multi-state model is a stochastic process $(X_t, t \in \mathcal{T})$ with a finite state space, where $X(t)$ represents the state occupied by the process at time t. In this paper we consider the progressive illness-death model depicted in Figure 1 and we assume that all subjects are in State 1 at time $t = 0$. Individuals may pass from the healthy state (State 1), to the disease state (State 2) and then to the absorbing dead state (State 3). Individuals are at risk of death in each transient state (States 1 and 2). This means that one individual may visit State 2 or go directly to State 3 without visiting State 2.

Let Z denote the sojourn time in State 1, T the total survival time of the process and $\rho = I(Z < T)$ the indicator of visiting State 2 at some time. For this model the transitions allowed are $1 \rightarrow 2$, $1 \rightarrow 3$ and

$2 \rightarrow 3$. As usual with survival data, individuals are generally followed over a certain period of time, providing right-censored observations. Let C denote the right-censoring variable, assumed to be independent of (Z, T) $(C \perp (Z, T))$. Due to censoring, rather than (Z, T) we observe $(\widetilde{Z}, \widetilde{T}, \Delta_1, \Delta_2)$ where $\widetilde{Z} = Z \wedge C$ and $\widetilde{T} = T \wedge C$ are the censored versions of Z and T, and $\Delta_1 = I(Z \leq C)$ and $\Delta = I(T \leq C)$ the respective censoring indicators. Let

$$\left\{ \left(\widetilde{Z}_i, \widetilde{T}_i, \Delta_{1i}, \Delta_i, \rho_i \right), 1 \leq i \leq n \right\}$$

be a random sample of the vector $\left(\widetilde{Z}, \widetilde{T}, \Delta_1, \Delta, \rho \right)$.

For two states h, j and two time points $s < t$, introduce the so-called transition probabilities

$$p_{hj}(s, t) = P(X(t) = j | X(s) = h).$$

In the illness-death model we have five different transition probabilities to estimate: $p_{11}(s, t)$, $p_{12}(s, t)$, $p_{13}(s, t)$, $p_{22}(s, t)$ and $p_{23}(s, t)$. Using the introduced notation, the transition probabilities can be written as

$$
\begin{aligned}
p_{11}(s, t) &= P\left(Z > t \mid Z > s\right), \\
p_{12}(s, t) &= P\left(Z \leq t, T > t \mid Z > s\right), \\
p_{13}(s, t) &= P\left(T \leq t \mid Z > s\right), \\
p_{22}(s, t) &= P\left(Z \leq t, T > t \mid Z \leq s, T > s\right), \\
p_{23}(s, t) &= P\left(T \leq t \mid Z \leq s, T > s\right).
\end{aligned}
$$

from which it follows

$$
\begin{aligned}
p_{11}(s, t) &= \frac{P\left(Z > t\right)}{P\left(Z > s\right)}, \\
p_{12}(s, t) &= \frac{P\left(s < Z \leq t, T > t\right)}{P(Z > s)}, \\
p_{13}(s, t) &= \frac{P\left(Z > s, T \leq t\right)}{P(Z > s)}, \\
p_{22}(s, t) &= \frac{P\left(Z \leq s, T > t\right)}{P\left(Z \leq s, T > s\right)}, \\
p_{23}(s, t) &= \frac{P\left(Z \leq s, s < T \leq t\right)}{P\left(Z \leq s, T > s\right)}.
\end{aligned}
$$

Since we have two obvious relations $p_{12}(s, t) = 1 - p_{11}(s, t) - p_{13}(s, t)$ and $p_{23}(s, t) = 1 - p_{22}(s, t)$ this means that in practice we only need to estimate three transition probabilities.

The inference in multi-state models is traditionally performed under the Markov assumption, which states that the relevant information for the future evolution of the process is provided by its current state,

independently of the states previously visited and the transition times among them. Under this assumption, the transition probabilities can be estimated nonparametrically using Aalen-Johansen estimators [9]. Their estimation method extends the time-honored Kaplan-Meier estimator to Markov chains. The Kaplan-Meier estimator is the standard method to estimate the survival function from time-to-event data that are subject to right censoring. It is a step function with jumps at event times. The size of the steps depends on the number of events and the number of individuals at risk at the corresponding time. Explicit formulae of the Aalen-Johansen estimators ($\mathtt{AJ}$) for the illness-death model are available [18]. In the following sections we introduce alternative non-Markov estimators for the same target.

2.2 Kaplan-Meier weighted estimators

For a general non-Markov illness-death process without recovery, [14] derived estimators for the transition probabilities defined in terms of multivariate "Kaplan-Meier integrals" with respect to the marginal distribution of the total time T. In particular, the estimators of $p_{12}(s, t)$ and $p_{22}(s, t)$ were proposed as an alternative to the Aalen-Johansen estimators in non-Markov situations. The transition probability $p_{11}(s, t)$ is defined as the ratio of observed survival distributions (and they can be estimated by the ordinary Kaplan-Meier estimator of survival [19] of the sojourn time in State 1, which we denote by $\widehat{S}_0$). The denominator of $p_{12}(s, t)$ can be estimated in the same way. The remaining quantities involve expectations of particular transformations of the pair (Z, T), $E\left[\varphi\left(Z, T\right)\right]$ which can not be estimated so simply

$$\widehat{p}_{12}^{\mathtt{LIDA}}(s, t) = \frac{\widehat{E}(\varphi_{s,t}(Z, T))}{\widehat{S}_0(s)} \tag{1}$$

and

$$\widehat{p}_{22}^{\mathtt{LIDA}}(s, t) = \frac{\widehat{E}(\widetilde{\varphi}_{s,t}(Z, T))}{\widehat{E}(\widetilde{\varphi}_{s,s}(Z, T))} \tag{2}$$

where $\varphi_{s,t}(u, v) = I(s < u \leq t, v > t)$ and $\widetilde{\varphi}_{s,t}(u, v) = I(u \leq s, v > t)$ and $\widehat{E}(\varphi_{s,t}(Z, T))$ is the "Kaplan-Meier integral"

$$\widehat{E}(\varphi_{s,t}(Z, T)) = \sum_i W_i \varphi_{s,t}(\widetilde{Z}_i, \widetilde{T}_i)$$

here W_i is the Kaplan-Meier weight attached to $\widetilde{T}_i$ when estimating the marginal distribution of T from

the $\left(\widetilde{T}_i, \Delta_i\right)$'s (equal to minus the jump at $\widetilde{T}_i$ of the Kaplan-Meier estimator of survival of the total time $\widehat{S}$). See [14] for more details.

The methods proposed by [14] have the drawback of requiring that the support of the censoring distribution contains the support of the lifetime distribution. An assumption that is often not fulfilled in medical applications due to limitations in the patient's following-up. To avoid this potential problem, corrected estimators were proposed by [16] for $p_{12}(s,t)$ and $p_{22}(s,t)$:

$$\widehat{p}_{12}^{\,\texttt{CLIDA}}(s,t) = \frac{\widehat{S}_0(s) - \widehat{S}_0(t) - \widehat{E}(\gamma_{s,t}(Z,T))}{\widehat{S}_0(s)} \quad (3)$$

and

$$\widehat{p}_{22}^{\,\texttt{CLIDA}}(s,t) = 1 - \frac{\widehat{E}(\widetilde{\gamma}_{s,t}(Z,T))}{\widehat{S}(t) - \widehat{S}_0(s)} \quad (4)$$

where $\gamma_{s,t}(u,v) = I(u > s, v \le t)$ and $\widetilde{\gamma}_{s,t}(u,v) = I(u \le s, s < v \le t)$.

All these quantities can be estimated nonparametrically using Kaplan-Meier weights.

2.3   Landmark Estimators

In this section we revisit the estimators proposed by [16] who propose alternative nonparametric estimators based on landmarking. The idea behind landmark estimation is to use the information given by the state occupied by the individual at a landmark time prior to time t to estimate the conditional probabilities $p_{ij}(s,t)$ (being s the landmark time). In practice, nonparametric estimators for the transition probabilities can be introduced by considering specific subsamples or portions of the data. In our case, to estimate the transition probabilities $p_{1j}(s,t)$, for $j = 1, 2, 3$, the analysis can be restricted to the individuals observed in State 1 at time s. As explained in [16], as long as C is independent of Z, a subject in $\mathcal{S}_1 = \left\{i : \widetilde{Z}_i > s\right\}$ 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 $\mathcal{S}_2 = \left\{i : \widetilde{Z}_i \le s < \widetilde{T}_i\right\}$, which serves to introduce landmark estimators for $p_{2j}(s,t)$, $j = 2, 3$. Accordingly, the following landmark estimators (LM) can be introduced,

$$\widehat{p}_{11}^{\,\texttt{LM}}(s,t) = \widehat{S}_0^{(s)}(t), \quad (5)$$

$$\widehat{p}_{13}^{\,\texttt{LM}}(s,t) = \widehat{S}^{(s)}(t), \quad (6)$$

where $\widehat{S}_0^{(s)}$ is the Kaplan-Meier estimator of survival of the sojourn time in State 1 computed in the subsample $\mathcal{S}_1$; and $\widehat{S}^{(s)}$ is the Kaplan-Meier estimator of survival of the total time computed in the same subsample. The formal relation $p_{12}(s,t) = 1 - p_{11}(s,t) - p_{13}(s,t)$ can be used to estimate p_{12}. Finally,

$$\widehat{p}_{23}^{\,\texttt{LM}}(s,t) = \widehat{S}^{[s]}(t), \quad (7)$$

where $\widehat{S}^{[s]}$ is the Kaplan-Meier estimator of survival of the total time computed in the subsample $\mathcal{S}_2$.

It is worth mention that [17] also introduced an estimator based on subsampling which they termed as a landmark Aalen-Johansen estimator of the transition probabilities. The idea behind the proposed estimator is to use the Aalen-Johansen estimator of the state occupation probabilities derived from those subsets (consisting of subjects occupying a given state at a particular time). Simulation studies published in the paper by [17] show that the landmark Aalen-Johansen estimator (LMAJ) and the landmark estimator (LM) perform similarly. In fact, the two landmark estimators (LM and LMAJ) of the transition probabilities $p_{11}(s,t)$, $p_{22}(s,t)$ and $p_{23}(s,t)$ are equivalent.

As a weakness, the landmark estimators [16] and [17] may provide large standard errors in estimation in some circumstances. This may occur for small sample sizes and/or large proportion of censored data. In such cases the estimators based on a landmark approach may result in a wiggly estimator with fewer jump points. A valid approach that can be used to reduce the variability of these estimators is to consider a modification of the landmark estimator based on presmoothing [20], [21].

2.4   Weighted Cumulative Hazard Estimators

In this section we propose new estimators for the transition probabilities $p_{11}(s,t)$, $p_{13}(s,t)$ and $p_{22}(s,t)$. The estimators are constructed using the cumulative hazard of the total time given a first time but where each observation has been weighted using the information of the first duration. The proposed estimator (WCH - weighted cumulative hazard) for the transition probability $p_{11}(s,t)$ is given by

$$\begin{aligned} \widehat{p}_{11}^{\,\texttt{WCH}}(s,t) &= \widehat{P}\left(Z > t \mid Z > s\right), \\ &= \prod_{v \in R_1} \{1 - \widehat{\Lambda}_{11}(dv)\}, \quad (8) \end{aligned}$$

where $\Lambda_{11}(dv)$ is the cumulative conditional hazard of Z given $Z > s$. Assuming that $Z \perp C$, $\Lambda_{11}(dv)$ can be estimated by

$$\widehat{\Lambda}_{11}(dv) = \frac{\sum_{i=1}^{n} I(\widetilde{Z}_i > s, \widetilde{Z}_i = v, \Delta_{1i} = 1)}{\sum_{i=1}^{n} I(\widetilde{Z}_i > s, \widetilde{Z}_i \geq v)}$$

and where $R_1 = \{\widetilde{Z}_i : \widetilde{Z}_i \leq t\}$.

Estimator (8) is equivalent to the estimator proposed by [14], [16] and the so-called Aalen-Johansen estimator [9].

Similar ideas can be used to obtain estimators for $p_{13}(s,t)$ and $p_{22}(s,t)$. Note that $p_{13}(s,t) = P(T \leq t \mid Z > s) = 1 - P(T > t \mid Z > s)$. Then,

$$\begin{aligned}
\widehat{p}_{13}^{\text{WCH}}(s,t) &= 1 - \widehat{P}(T > t \mid Z > s), \\
&= 1 - \prod_{v \in R_{13}} \{1 - \widehat{\Lambda}_{13}(dv)\} \prod_{v \in R_{123}} \{1 - \widehat{\Lambda}_{123}(dv)\}, \quad (9)
\end{aligned}$$

where $\Lambda_{13}(dv)$ is the cumulative conditional hazard of T given $Z > s$ for those individuals going directly into State 3 without visiting State 2; and $\Lambda_{123}(dv)$ is the cumulative conditional hazard of T given $Z > s$ for those visiting State 2. Assuming that $(Z,T) \perp C$, $\Lambda_{13}(dv)$ can be estimated by

$$\widehat{\Lambda}_{13}(dv) = \frac{\sum_{i=1}^{n} I(\widetilde{Z}_i > s, \widetilde{Z}_i = \widetilde{T}_i, \widetilde{T}_i = v, \Delta_{2i} = 1)}{\sum_{i=1}^{n} I(\widetilde{Z}_i > s, \widetilde{T}_i \geq v)}$$

and where $R_{13} = \{\widetilde{T}_i : \widetilde{T}_i \leq t\}$; whereas $\Lambda_{123}(dv)$ can be estimated by

$$\widehat{\Lambda}_{123}(dv) = \frac{\sum_{i=1}^{n} I(\widetilde{Z}_i > s, \widetilde{Z}_i < \widetilde{T}_i, \widetilde{T}_i = v, \Delta_{2i} = 1)}{\sum_{i=1}^{n} I(\widetilde{Z}_i > s, \widetilde{T}_i \geq v)}$$

and where $R_{123} = \{\widetilde{T}_i : \widetilde{T}_i \leq t\}$.

Then, $p_{12}(s,t)$ can be estimated by $\widehat{p}_{12}^{\text{WCH}}(s,t) = 1 - \widehat{p}_{11}^{\text{WCH}}(s,t) - \widehat{p}_{13}^{\text{WCH}}(s,t)$. Note that estimators of $\widehat{p}_{1j}^{\text{WCH}}(s,t)$, $j = 1, 2$ are equivalent to the landmark estimators proposed by [16].

Since $p_{22}(s,t) = \frac{P(Z<s,T>t)}{P(Z<s,T>s)} = \frac{P(T>t|Z<s)}{P(T>s|Z<s)}$. The key to estimating $p_{22}(s,t)$ is to estimate $P(T > u \mid Z < s)$ for $u \in \{s,t\}$. These quantities can be estimated by

$$\widetilde{P}(T > u \mid Z < s) = \prod_{v \in R_{23}} \{1 - \widetilde{\Lambda}_{23}(dv)\} \quad (10)$$

where $R_{23} = \{\widetilde{T}_{23i} : \widetilde{T}_{23i} \leq u - \widetilde{Z}_i, \widetilde{Z}_i < \widetilde{T}_i\}$; and $\Lambda_{23}(dv)$ can be estimated by

$$\widetilde{\Lambda}_{23}(\Delta v) = \frac{\sum_{i=1}^{n} I(\widetilde{Z}_i \leq s, \widetilde{Z}_i < \widetilde{T}_i, \widetilde{T}_{23i} = v, \Delta_{2i} = 1)/\widehat{G}(\widetilde{Z}_i + v)}{\sum_{i=1}^{n} I(\widetilde{Z}_i \leq s, \widetilde{Z}_i < \widetilde{T}_i, \widetilde{T}_{23i} \geq v, \Delta_{1i} = 1)/\widehat{G}(\widetilde{Z}_i + v)}.$$

The resultant estimator is labeled as $\widehat{p}_{22}^{\text{WCH}}(s,t)$.

Since $p_{22}(s,t) = P(T > t|Z < s, T > s)$, an alternative estimator is given by

$$\widetilde{p}_{22}^{\text{WCH}}(s,t) = \widetilde{P}(T > u \mid Z < s, T > s) = \prod_{v \in R_{23}^{\star}} \{1 - \widetilde{\Lambda}_{23}^{\star}(dv)\}, (11)$$

where $R_{23}^{\star} = \{\widetilde{T}_i : \widetilde{T}_i \leq t\}$ and where $\Lambda_{23}^{\star}(dv)$ can be estimated by

$$\widetilde{\Lambda}_{23}^{\star}(\Delta v) = \frac{\sum_{i=1}^{n} I(\widetilde{Z}_i \leq s, \widetilde{T}_i > s, \widetilde{T}_i = v, \Delta_{2i} = 1)/\widehat{G}(v)}{\sum_{i=1}^{n} I(\widetilde{Z}_i \leq s, \widetilde{T}_i > s, \widetilde{T}_i \geq v, \Delta_{1i} = 1)/\widehat{G}(\max(\widetilde{Z}_i, v))}.$$

The estimator $\widetilde{p}_{22}^{\text{WCH}}(s,t)$ is equivalent to the landmark estimator $\widehat{p}_{22}^{\text{LM}}(s,t)$ proposed by [16].

The estimation of the variance is important for inference purposes. Resampling techniques such as the bootstrap provides here a practical solution to the problem of variance estimation and inference. These methods can be used to construct confidence limits based on the percentile bootstrap.

3 Simulation Study

In this section we investigate the performance of the proposed estimators through simulations. More specifically, the estimators introduced in Section 2 are considered. In particular we aim to compare the performance of the Aalen-Johansen estimator which benefits from the assumption of Markovianity on the underlying stochastic process, with alternative estimators which are free of the Markov condition. The simulation addresses also the question about the more efficient estimator in different scenarios.

To simulate the data in the irreversible illness-death model, we separately consider the subjects passing through State 2 at some time (that is, those cases with $\rho = 1$), and those who directly go to the absorbing State 3 ($\rho = 0$). For the second subgroup of individuals ($\rho = 0$), times to death without illness are generated from the hazard function $h_{13}(t) = 0.024 \times t$. For the first subgroup of individuals ($\rho = 1$), the successive gap times $(Z, T - Z)$ are simulated using two cause-specific hazard functions, h_{12} and h_{23} for each of the events (illness and death). The cause-specific hazard for the intermediate event was defined as $h_{12}(t) = \frac{0.29}{t+1}$. For the individuals that experienced the disease, times to death after the disease were generated using three different hazards:

$$h_{23}^1(t, t_{12}) = 0.05$$
$$h_{23}^2(t, t_{12}) = \frac{1}{0.25(t_{12} + 1)^{0.8}}$$
$$h_{23}^3(t, t_{12}) = 0.04 \times log(t + 1)$$

where $t > 0$, denotes the time since the start point, and t_{12} is the transition time from State 1 to State 2.

The use of these three different hazard functions provides three different scenarios. The first scenario can be considered Markovian since the hazard of death after the disease was set constant being independent of t and t_{12}. In the two remaining scenarios, the hazard for death after disease depended on the these times. The second scenario is semi-Markovian since the process depend not only the current state, but also how long it has been in the current state (time refers to time since entering the intermediate state). The third scenario is non-Markov since the hazard for death after disease depended on the time since entry in study.

An independent uniform censoring time C is generated, according to models $U[0, 40]$ and $U[0, 30]$. For the Markovian scenario, the first model, presents 19% of censoring on the first gap time Z, 40% censoring on the total time T and 42% on the second gap time $T - Z$, for those individuals who entered in state 2. The second model changes these censoring levels to 33%, 58% and about 36%, respectively. The first model in the semi-Markovian scenario, reveals 23% of censoring for Z, 40% for T and 41% for $T - Z$. Second model increases these censoring levels to 36%, 49% and 41%, respectively. Finally, in the non-Markovian scenario, the model $U[0, 40]$ provides 25% of censoring on the first gap time, 43% on the total time and 41% on the second gap time. Under the model $U[0, 30]$ censoring increases to 26%, 44% and about 42%, respectively.

For each simulated scenario we consider several different points (s, t) pairs, corresponding to combinations of times 2, 4, 8 and 12 representing the differences between closer and distant times. Sample sizes $n = 100$ and $n = 250$ are considered. In each simulation, 1000 samples are generated. From these samples we obtained the mean for all generated data sets. As a measure of efficiency, we took the Mean Squared Error (MSE) but we also computed the standard deviations (SD) and the Bias.

Tables 1 (Markov scenario), 2 (semi-Markov scenario) and 3 (non-Markov scenario) report the results for transition probability $p_{13}(s, t)$. When one is confident of the Markov assumption, the Aalen-Johansen is preferred over non-Markovian estimators since it

agreeing with results reported in Table 1. Results reported in the Tables 2 and 3 also reveal that the Aalen-Johansen estimator (labeled as AJ) might still perform reports a smaller variance in estimation. This is in reasonably well in situations where the process shows only mild deviations from Markovianity. However, when there is strong evidence that the process is not Markov the use of a non-Markov estimator is preferable due to their greater accuracy. This can be observed from results reported in Tables 2 and 3.

All three simulation scenarios reveal that the performance of the methods is poorer at the right tail. This was expected because for larger values of s and t, the censoring effects are stronger. The SD decreases with an increase of the sample size and with the decrease of the censoring percentage, which was also expected.

Results in Tables 1, 2 and 3 reveal a poor performance of the original non-Markov estimators by [14], referred to as LIDA estimators. The two methods alternative non-Markov methods (the corrected LIDA, labeled as cLIDA, and the Weighted Cumulative Hazard method WCH) obtain in all settings a negligible bias (decreasing as the sample size increases), while the LIDA estimator show a systematic bias.

Tables 2 and 3 show that the Markov-free estimators cLIDA and WCH may behave much more efficiently than the Aalen-Johansen. This is because of the failure of the Markov assumption from which the Aalen-Johansen estimator is built. This is more evident in the semi-Markov scenario, with higher lag times $t - s$. In these cases, the Aalen-Johansen show a systematic bias which does not decrease with an increasing sample size. In these cases the application of the Aalen-Johansen method is not recommended here, due to possible biases. The poor behavior of the Aalen-Johansen estimator can also be seen in Figure 2, in which we show the boxplots of the estimates of the transition probabilities based on the 1000 Monte Carlo replicates for the four estimators, with different sample sizes. From these plots it can be seen that the cLIDA and WCH methods are unbiased estimators and confirm the less variability of the Aalen-Johansen estimator. The WCH method (which in this case is equivalent to the LM method) is the preferred since is the unbiased method reporting less variability.

Tables 4, 5 and 6 report the results for five different estimators for the transition probability $p_{22}(s, t)$. Results reported in Table 4 reveal that the Aalen-Johansen estimator is the preferred since it reports unbiased estimates with smaller variance in estimation. This was expected since the process is Markovian in this scenario. Again, it is important mentioning that this estimator which assumes the process to be Markovian still perform reasonably well in situations where

Table 1: Bias and standard deviation (SD) for the three estimators of $p_{13}(s, t)$. Markov scenario with two sample sizes and two censoring levels.

(s,t)=		$\hat{p}_{13}^{AJ}(s,t)$ bias	SD	$\hat{p}_{13}^{LIDA}(s,t)$ bias	SD	$\hat{p}_{13}^{CLIDA}(s,t)$ bias	SD	$\hat{p}_{13}^{MCH}(s,t)$ bias	SD
(2,4)									
n=100	$C \sim U[0,30]$	<0.0001	00299	0.0345	0.0533	<0.0001	0.0316	<0.0001	0.0316
	$C \sim U[0,20]$	<0.0001	0.0300	0.0600	0.0552	<0.0001	0.0312	<0.0001	0.0311
n=250	$C \sim U[0,30]$	<0.0001	0.0187	0.0335	0.0368	<0.0001	0.0197	<0.0001	0.0197
	$C \sim U[0,20]$	<0.0001	0.0193	0.0584	0.0383	<0.0001	0.0204	<0.0001	0.0204
(2,8)									
n=100	$C \sim U[0,30]$	0.0024	0.0527	0.0659	0.0926	0.0021	0.0567	0.0025	0.0563
	$C \sim U[0,20]$	0.0014	0.0594	0.118	0.0907	0.0010	0.0630	<0.0001	0.0627
n=250	$C \sim U[0,30]$	<0.0001	0.0340	0.0677	0.0626	<0.0001	0.0363	<0.0001	0.036
	$C \sim U[0,20]$	<0.0001	0.0375	0.1147	0.0607	<0.0001	0.0406	<0.0001	0.0401
(4,12)									
n=100	$C \sim U[0,30]$	-0.0039	0.0736	0.0669	0.1034	-0.0029	0.0791	-0.0038	0.0767
	$C \sim U[0,20]$	-0.0024	0.0824	0.1084	0.1181	-0.0013	0.0911	-0.0014	0.0893
n=250	$C \sim U[0,30]$	0.0011	0.0462	0.0671	0.0733	0.0018	0.0508	0.0017	0.0496
	$C \sim U[0,20]$	<0.0001	0.0503	0.1132	0.0708	<0.0001	0.0541	<0.0001	0.0533
(8,12)									
n=100	$C \sim U[0,30]$	-0.0015	0.0816	0.0424	0.103	-0.0031	0.0844	-0.0029	0.0839
	$C \sim U[0,20]$	-0.0019	0.099	0.0645	0.117	-0.0016	0.1034	-0.0027	0.1012
n=250	$C \sim U[0,30]$	<0.0001	0.0529	0.0412	0.0658	<0.0001	0.0551	<0.0001	0.0546
	$C \sim U[0,20]$	<0.0001	0.0621	0.0615	0.0785	<0.0001	0.0652	<0.0001.	0.0639

Table 2: Bias and standard deviation (SD) for the three estimators of $p_{13}(s, t)$. Semi-Markov scenario with two sample sizes and two censoring levels.

(s,t)=		$\hat{p}_{13}^{AJ}(s,t)$ bias	SD	$\hat{p}_{13}^{LIDA}(s,t)$ bias	SD	$\hat{p}_{13}^{CLIDA}(s,t)$ bias	SD	$\hat{p}_{13}^{MCH}(s,t)$ bias	SD
(2,4)									
n=100	$C \sim U[0,30]$	0.0085	0.0312	0.0181	0.0488	0.0015	0.0323	0.0016	0.0323
	$C \sim U[0,20]$	0.0057	0.0311	0.0333	0.0534	<0.0001	0.0331	<0.0001	0.0331
n=250	$C \sim U[0,30]$	0.0071	0.0197	0.0158	0.0337	<0.0001	0.0202	<0.0001	0.0202
	$C \sim U[0,20]$	0.0065	0.0201	0.0333	0.0378	<0.0001	0.0208	<0.0001	0.0208
(2,8)									
n=100	$C \sim U[0,30]$	0.0273	0.0579	0.0498	0.087	0.0014	0.0617	0.0014	0.0610
	$C \sim U[0,20]$	0.0254	0.0596	0.0863	0.0932	<0.0001	0.0639	<0.0001	0.0634
n=250	$C \sim U[0,30]$	0.0276	0.0353	0.0474	0.0590	0.0019	0.0369	0.0022	0.0370
	$C \sim U[0,20]$	0.0257	0.0354	0.0840	0.0629	<0.0001	0.0383	<0.0001	0.0378
(4,12)									
n=100	$C \sim U[0,30]$	0.0328	0.0757	0.0761	0.1059	0.0035	0.0798	0.0035	0.0783
	$C \sim U[0,20]$	0.0308	0.0841	0.1193	0.1079	0.0028	0.0902	0.0022	0.0883
n=250	$C \sim U[0,30]$	0.0270	0.0459	0.0672	0.0727	-0.0024	0.0498	-0.0024	0.0486
	$C \sim U[0,20]$	0.0271	0.0530	0.1106	0.0744	-0.0027	0.0573	-0.0020	0.0562
(8,12)									
n=100	$C \sim U[0,30]$	0.0085	0.0842	0.0505	0.1056	<0.0001	0.0872	-0.0012	0.0861
	$C \sim U[0,20]$	0.0077	0.0987	0.0738	0.1216	0.0013	0.1040	-0.0013	0.1004
n=250	$C \sim U[0,30]$	0.0095	0.0530	0.0508	0.0673	<0.0001	0.0550	<0.0001	0.0542
	$C \sim U[0,20]$	0.0085	0.0596	0.0720	0.0750	-0.0010	0.0620	-0.0011	0.0608

Table 3: Bias and standard deviation (SD) for the three estimators of $p_{13}(s,t)$. Non-Markov scenario with three sample sizes and two censoring levels.

| | | $\hat{p}_{13}^{\text{AJ}}(s,t)$ | | $\hat{p}_{13}^{\text{LIDA}}(s,t)$ | | $\hat{p}_{13}^{\text{cLIDA}}(s,t)$ | | $\hat{p}_{13}^{\text{WCH}}(s,t)$ | |
		bias	SD	bias	SD	bias	SD	bias	SD
(s,t)=	(2,4)								
n=100	$C \sim U[0,30]$	0.0018	0.0282	0.0147	0.0488	-0.0019	0.0286	-0.0019	0.0285
	$C \sim U[0,20]$	0.0035	0.0318	0.0376	0.0563	<0.0001	0.0325	<0.0001	0.0325
n=250	$C \sim U[0,30]$	0.0034	0.0185	0.0123	0.0345	<0.0001	0.0188	<0.0001	0.0188
	$C \sim U[0,20]$	0.0030	0.0191	0.0309	0.0413	<0.0001	0.0194	<0.0001	0.0194
(s,t)=	(2,8)								
n=100	$C \sim U[0,30]$	0.0131	0.0563	0.0252	0.0936	-0.0017	0.0588	-0.0014	0.0586
	$C \sim U[0,20]$	0.0119	0.0567	0.0792	0.0925	<0.0001	0.0603	<0.0001	0.0596
n=250	$C \sim U[0,30]$	0.0164	0.0349	0.0257	0.0584	0.0027	0.0367	0.0028	0.0364
	$C \sim U[0,20]$	0.0140	0.0365	0.0686	0.0711	<0.0001	0.0392	<0.0001	0.0390
(s,t)=	(4,12)								
n=100	$C \sim U[0,30]$	0.0176	0.0749	0.0316	0.1103	<0.0001	0.0801	-0.0015	0.0786
	$C \sim U[0,20]$	0.0212	0.0832	0.0900	0.1197	0.0045	0.0906	0.0034	0.0869
n=250	$C \sim U[0,30]$	0.0204	0.0468	0.0240	0.0783	<0.0001	0.0506	<0.0001	0.0498
	$C \sim U[0,20]$	0.0179	0.0527	0.0815	0.0841	-0.0017	0.0583	-0.0015	0.0567
(s,t)=	(8,12)								
n=100	$C \sim U[0,30]$	0.0145	0.0814	0.0259	0.1077	0.0039	0.0842	0.0038	0.0835
	$C \sim U[0,20]$	0.0060	0.1005	0.0610	0.1246	-0.0038	0.1032	-0.0034	0.1031
n=250	$C \sim U[0,30]$	0.0087	0.0527	0.0189	0.0709	<0.0001	0.0541	<0.0001	0.0539
	$C \sim U[0,20]$	0.0114	0.0619	0.0596	0.0815	0.0016	0.0645	0.0017	0.0636
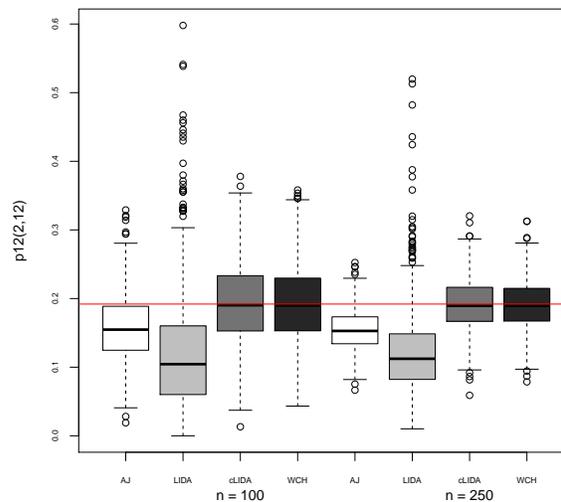
Figure 2: Boxplots of the M = 1000 estimates of the transition probabilities of the $\hat{p}_{12}^{\text{AJ}}$, $\hat{p}_{12}^{\text{LIDA}}$, $\hat{p}_{12}^{\text{cLIDA}}$ and $\hat{p}_{12}^{\text{WCH}}$ with two different samples sizes for semi-Markovian scenario. Censoring times were generated from an uniform distribution on [0, 30].

the process shows only mild deviations from Marko-vianity. This occurs for example in the semi-Markov scenario with small lag times $t - s$. In these cases, the Aalen-Johansen reports estimates with small bias but less variability and therefore low mean squared errors. As the lag times $t - s$ increase so the bias resulting in a clear biased estimator. This behavior is also present in the non-Markov scenario (Table 6). Results shown in Tables 5 and 6 reveals that when there is strong evidence that the process is not Markov that the use of a non-Markov estimator is preferable. With the exception of the LIDA method all the remaining non-Markov methods (cLIDA, LM and WCH) are valid alternative estimators due to their greater accuracy. Again, the performance of the LIDA method is poorer even worst than the Aalen-Johansen estimator. Simulation results reveal that the LIDA estimator is systematically (downward) biased whereas the three non-Markov methods cLIDA, LM and WCH are asymptotically unbiased. The best performance is attained by the non-Markov methods (cLIDA, LM and WCH) which lead to more efficient estimation of the transition probabilities. This can be seen in all measures (bias, standard deviation and mean square error). However, when considering all scenarios and all pairs (s,t) neither of the two methods seems to be uniformly best for estimating $p_{22}(s,t)$. However,

the landmark method `LM` reveals in most cases less variability and therefore better results (with less mean square errors) than the remaining non-Markov estimators.

For completeness purposes we show in Figures 3, 4 and 5 the boxplots of the estimates of the transition probability $p_{22}(s,t)$ based on 1000 Monte Carlo replicates for the five estimators, with different sample sizes. The boxplots shown in these figures are in agree with our findings reported in Table 4, 5 and 6. From these plots it can be seen that the `LIDA` estimator of $p_{23}(s,t)$ is systematically (downward) biased and that the `AJ` estimator may also lead to biased estimates (but with less variability) under deviations from Markovianity. Under a Markov scenario (Figure 3), all estimators but the LIDA estimator revealed to be unbiased and with a variance that decrease with the sample size. The `AJ` estimator is preferable in this case because it provides less variability. When the multi-state model is not Markov, this is no longer the case. Despite of offering a small variability, the bias associated to Aalen-Johansen estimator in non-Markov scenarios (Figures 4 and 5) makes this approach unappropriated. The methods labeled as `LM` and `WCH` are recommended in these cases.
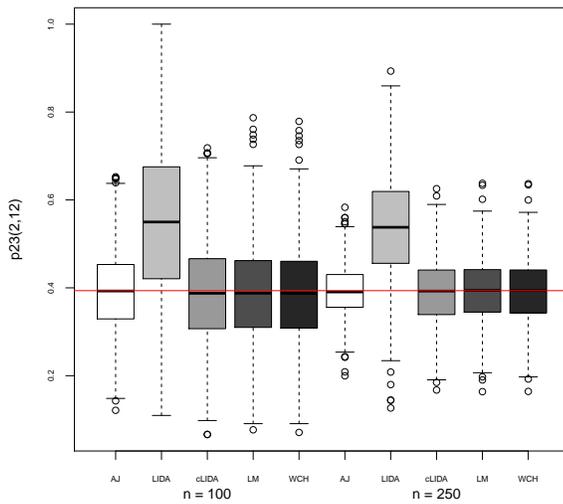

Figure 3: Boxplots of the M = 1000 estimates of the transition probabilities of the $\hat{p}_{22}^{AJ}$, $\hat{p}_{22}^{LIDA}$, $\hat{p}_{22}^{cLIDA}$, $\hat{p}_{22}^{LM}$ and $\hat{p}_{22}^{WCH}$ with two different samples sizes for Markovian scenario. Censoring times were generated from an uniform distribution on [0, 30].

Table 4: Bias and standard deviation (SD) for the five estimators of $p_{22}(s, t)$. Markov scenario with two sample sizes and two censoring levels.

(s,t)=		$\hat{p}_{22}^{AJ}(s,t)$ bias	SD	$\hat{p}_{22}^{LIDA}(s,t)$ bias	SD	$\hat{p}_{22}^{cLIDA}(s,t)$ bias	SD	$\hat{p}_{22}^{LM}(s,t)$ bias	SD	$\hat{p}_{22}^{WCH}(s,t)$ bias	SD
	(2,4)										
n=100	$C \sim U[0,30]$	0.0017	0.0547	-0.0409	0.0960	<0.0001	0.0597	<0.0001	0.0599	<0.0001	0.0598
	$C \sim U[0,20]$	<0.0001	0.0547	-0.0926	0.1330	-0.0012	0.0598	-0.0014	0.0596	-0.0013	0.0596
n=250	$C \sim U[0,30]$	0.0010	0.0341	-0.0361	0.0552	<0.0001	0.0371	<0.0001	0.0372	<0.0001	0.0372
	$C \sim U[0,20]$	-0.0012	0.0365	-0.0778	0.0818	<0.0001	0.0410	<0.0001	0.0408	<0.0001	0.0408
(s,t)=	(2,8)										
n=100	$C \sim U[0,30]$	<0.0001	0.0788	-0.1129	0.1626	<0.0001	0.0969	0.0013	0.0962	0.0013	0.0963
	$C \sim U[0,20]$	<0.0001	0.0812	-0.2364	0.2174	-0.0011	0.1017	-0.0013	0.0993	<0.0001	0.0992
n=250	$C \sim U[0,30]$	0.0030	0.0488	-0.0949	0.0976	0.0033	0.0597	0.0028	0.0592	0.0029	0.0592
	$C \sim U[0,20]$	-0.0029	0.0526	-0.2143	0.1363	-0.0025	0.0651	-0.0029	0.0634	-0.0028	0.0635
(s,t)=	(4,12)										
n=100	$C \sim U[0,30]$	-0.0015	0.0845	-0.1550	0.1687	<0.0001	0.0981	-0.0016	0.0967	-0.0016	0.0972
	$C \sim U[0,20]$	0.0018	0.0958	-0.3451	0.2312	0.0026	0.1119	0.0028	0.1078	0.0017	0.1081
n=250	$C \sim U[0,30]$	<0.0001	0.0555	-0.1376	0.1055	0.0021	0.0635	0.0019	0.0622	0.0017	0.0627
	$C \sim U[0,20]$	-0.0021	0.0589	-0.2968	0.1566	-0.0023	0.0689	-0.0017	0.0668	-0.0019	0.0675
(s,t)=	(8,12)										
n=100	$C \sim U[0,30]$	<0.0001	0.0789	-0.1262	0.1568	-0.0026	0.0821	-0.0028	0.0819	-0.0030	0.0830
	$C \sim U[0,20]$	-0.0046	0.0952	-0.3220	0.2814	-0.0056	0.0992	-0.0065	0.0979	-0.0084	0.1024
n=250	$C \sim U[0,30]$	-0.0018	0.0497	-0.1030	0.0930	-0.0019	0.0513	-0.0018	0.0507	-0.0018	0.0514
	$C \sim U[0,20]$	<0.0001	0.0578	-0.2594	0.1730	<0.0001	0.0615	<0.0001	0.0605	<0.0001	0.0618

Table 5: Bias and standard deviation (SD) for the five estimators of $p_{22}(s,t)$. Semi-Markov scenario with two sample sizes and two censoring levels.

(s,t)=	$\hat{p}_{22}^{AJ}(s,t)$ bias	SD	$\hat{p}_{22}^{LIDA}(s,t)$ bias	SD	$\hat{p}_{22}^{CLLIDA}(s,t)$ bias	SD	$\hat{p}_{22}^{LM}(s,t)$ bias	SD	$\hat{p}_{22}^{WCH}(s,t)$ bias	SD
(2,4)										
n=100										
$C \sim U[0,30]$	0.0207	0.0842	-0.0105	0.1081	-0.0013	0.0989	<0.0001	0.0976	<0.0001	0.098
$C \sim U[0,20]$	0.0191	0.0908	-0.0333	0.1266	<0.0001	0.1047	<0.0001	0.1040	<0.0001	0.1048
n=250										
$C \sim U[0,30]$	0.0230	0.0538	-0.0039	0.0668	0.0027	0.0625	0.0033	0.0614	0.0034	0.0620
$C \sim U[0,20]$	0.0186	0.0576	-0.0279	0.0813	-0.0038	0.0677	-0.0037	0.0664	-0.0036	0.0668
(2,8)										
n=100										
$C \sim U[0,30]$	0.0805	0.0901	-0.0182	0.1388	<0.0001	0.1280	0.0022	0.1207	0.0027	0.1208
$C \sim U[0,20]$	0.0776	0.0948	-0.0704	0.1606	-0.0011	0.1318	0.0000	0.1225	<0.0001	0.1250
n=250										
$C \sim U[0,30]$	0.0799	0.0544	-0.0122	0.0858	0.0025	0.0749	0.0042	0.0704	0.0041	0.0712
$C \sim U[0,20]$	0.0794	0.0590	-0.0524	0.1075	<0.0001	0.0839	-0.0010	0.0761	-0.0010	0.0773
(4,12)										
n=100										
$C \sim U[0,30]$	0.0665	0.1004	-0.0626	0.1498	-0.0017	0.1269	-0.0024	0.1179	-0.0034	0.1216
$C \sim U[0,20]$	0.0732	0.1115	-0.1436	0.1847	0.004	0.1481	0.0058	0.1337	<0.0001	0.1414
n=250										
$C \sim U[0,30]$	0.0703	0.0612	-0.0442	0.1000	0.0026	0.0822	0.0026	0.0736	0.0027	0.0759
$C \sim U[0,20]$	0.0714	0.0691	-0.1196	0.1285	0.0011	0.0941	0.0043	0.0822	0.0049	0.0860
(8,12)										
n=100										
$C \sim U[0,30]$	0.0207	0.1201	-0.0949	0.1845	0.0017	0.1376	0.0027	0.1308	<0.0001	0.1382
$C \sim U[0,20]$	0.0138	0.1381	-0.2551	0.2722	-0.0037	0.1572	-0.0056	0.1484	-0.0136	0.1712
n=250										
$C \sim U[0,30]$	0.0189	0.0742	-0.0693	0.1117	0.0018	0.0825	0.0027	0.0791	0.0028	0.0823
$C \sim U[0,20]$	0.0197	0.0893	-0.1973	0.1748	0.0016	0.1031	0.0025	0.0962	-0.0016	0.1052

Table 6: Bias and standard deviation (SD) for the five estimators of $p_{22}(s,t)$. Non-Markov scenario with three sample sizes and two censoring levels.

(s,t)=	$\hat{p}_{22}^{AJ}(s,t)$ bias	SD	$\hat{p}_{22}^{LIDA}(s,t)$ bias	SD	$\hat{p}_{22}^{CLLIDA}(s,t)$ bias	SD	$\hat{p}_{22}^{LM}(s,t)$ bias	SD	$\hat{p}_{22}^{WCH}(s,t)$ bias	SD
(2,4)										
n=100										
$C \sim U[0,30]$	0.0093	0.0492	-0.0074	0.0661	<0.0001	0.0573	<0.0001	0.0571	<0.0001	0.0576
$C \sim U[0,20]$	0.0058	0.0512	-0.0322	0.0857	-0.0016	0.0586	-0.0018	0.0584	-0.0022	0.0588
n=250										
$C \sim U[0,30]$	0.0071	0.0308	-0.0078	0.0408	-0.0011	0.0362	-0.0011	0.0360	-0.0014	0.0362
$C \sim U[0,20]$	0.0082	0.0318	-0.0262	0.0513	<0.0001	0.0373	<0.0001	0.0370	<0.0001	0.0372
(2,8)										
n=100										
$C \sim U[0,30]$	0.0347	0.0797	-0.0300	0.1260	-0.0026	0.1000	-0.0017	0.0992	-0.0023	0.0996
$C \sim U[0,20]$	0.0334	0.0843	-0.1050	0.1680	0.0033	0.1058	0.0017	0.1050	0.0014	0.1051
n=250										
$C \sim U[0,30]$	0.0309	0.0490	-0.0268	0.0781	-0.0030	0.0629	-0.0028	0.0621	-0.0035	0.0624
$C \sim U[0,20]$	0.0307	0.0523	-0.0992	0.1037	-0.0046	0.0663	-0.0045	0.0642	-0.0052	0.0646
(4,12)										
n=100										
$C \sim U[0,30]$	0.0323	0.0890	-0.0511	0.1340	0.0015	0.1065	<0.0001	0.1026	-0.0025	0.1042
$C \sim U[0,20]$	0.0325	0.1027	-0.1729	0.1922	0.0019	0.1259	0.0021	0.1213	-0.0015	0.1240
n=250										
$C \sim U[0,30]$	0.0322	0.0552	-0.0410	0.0871	<0.0001	0.0668	-0.0010	0.0643	-0.0036	0.0648
$C \sim U[0,20]$	0.0387	0.0626	-0.1504	0.1286	0.0064	0.0793	0.0068	0.0746	0.0041	0.0760
(8,12)										
n=100										
$C \sim U[0,30]$	0.0085	0.0934	-0.0463	0.1270	-0.0035	0.1015	-0.0040	0.0999	-0.0097	0.1030
$C \sim U[0,20]$	0.0145	0.1061	-0.1854	0.2219	0.0032	0.1157	0.0041	0.1125	-0.0046	0.1220
n=250										
$C \sim U[0,30]$	0.0115	0.0575	-0.0340	0.0783	<0.0001	0.0619	<0.0001	0.0606	-0.0054	0.0639
$C \sim U[0,20]$	0.0111	0.0689	-0.1534	0.1366	<0.0001	0.0740	<0.0001	0.0727	-0.0067	0.0766

Figure 4: Boxplots of the M = 1000 estimates of the transition probabilities of the $\widehat{p}_{22}^{\texttt{AJ}}$, $\widehat{p}_{22}^{\texttt{LIDA}}$, $\widehat{p}_{22}^{\texttt{cLIDA}}$, $\widehat{p}_{22}^{\texttt{LM}}$ and $\widehat{p}_{22}^{\texttt{WCH}}$ with two different samples sizes for semi-Markovian scenario. Censoring times were generated from an uniform distribution on [0, 30].

Figure 5: Boxplots of the M = 1000 estimates of the transition probabilities of the $\widehat{p}_{22}^{\texttt{AJ}}$, $\widehat{p}_{22}^{\texttt{LIDA}}$, $\widehat{p}_{22}^{\texttt{cLIDA}}$, $\widehat{p}_{22}^{\texttt{LM}}$ and $\widehat{p}_{22}^{\texttt{WCH}}$ with two different samples sizes for non-Markovian scenario. Censoring times were generated from an uniform distribution on [0, 30].

4 Example of Application

Colorectal cancer is one of the most commonly diagnosed cancers worldwide. It is also the one of most frequent causes of cancer-related death in both men and women. Several lifestyle-related factors have been linked to colorectal cancer including diet, weight and exercise. Survival rates for colorectal cancer vary worldwide but they have been associated to several clinical and pathological factors including age, tumor size, lymph nodes with detectable cancer, etc.

Surgical resection is the best treatment option for patients with colorectal cancer and the most powerful tool for assessing prognosis following potentially curative surgery. In a large percentage of the patients with colorectal cancer, the diagnosis is made at a sufficiently early stage when all apparent disease tissue can be surgically removed. Unfortunately, some of these patients have residual cancer, which leads to recurrence of the disease and death (in some cases). Cancer patients who have experienced a recurrence are known to be at a substantially higher risk of mortality. This mortality is higher in cases of early recurrences. Traditionally, the effect of these covariates is studied using the Cox proportional hazards model [22] with time-dependent covariates. The analysis of such studies can also be successfully performed using a multi-state model [3, 8].

In this section we re-analyzed data from one of the first successful trials of adjuvant chemotherapy for colon cancer [23]. In this data set we have a total of 929 patients from a large clinical trial on Duke's stage III patients, affected by colon cancer, that underwent a curative surgery for colorectal cancer. In this study, patients were followed from the date of cancer diagnosis until censoring or death. A total of 468 patients developed a recurrence and among these 414 died; 38 patients died without recurrence. The rest of the patients (423) remained alive and disease-free up to the end of the follow-up. Cancer recurrence affects the patient's outcome and can be included as a transient state in a progressive illness-death model with states "alive and disease-free" (State 1), "alive with recurrence" (State 2) and "death" (State 3).

Besides recurrence, the sojourn time in State 1, the total time of the process and the corresponding indicator statuses are known for each individual. Additional covariates such as age, sex, tumor size and lymph nodes with detectable cancer, are also available. In this paper we consider early recurrence as recurrence within 1 year after primary surgery of colorectal cancer. Considering overall survival after recurrence, the median survival time in the early recurrence group was 498 days with overall survival rates of 26.6%, 10.8% and 6.3% for 2, 3 and 5 years after
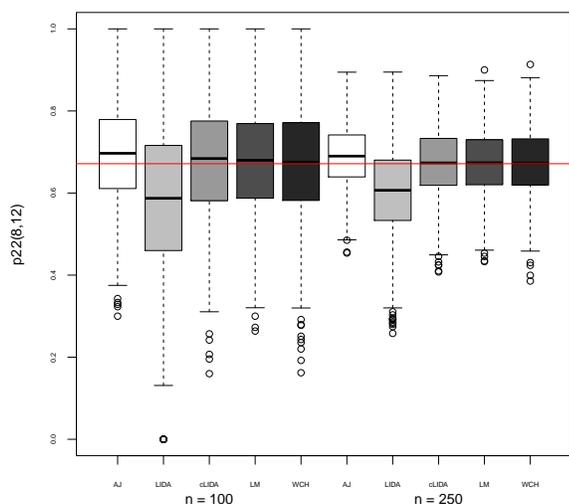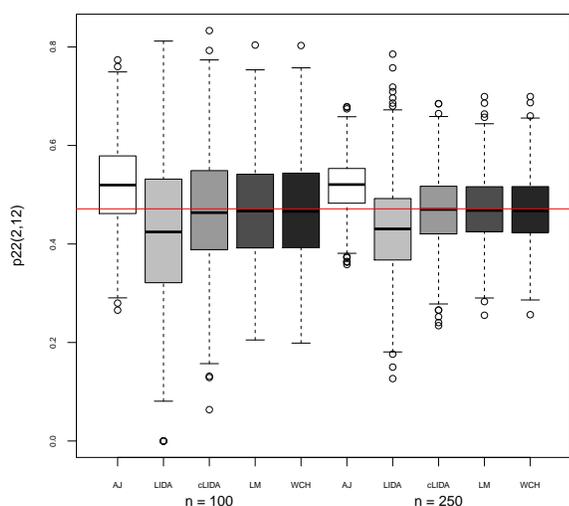
surgery. As expected, better results (P-value < 0.001) were obtained for patients in the late recurrence group (i.e., with a time to recurrence greater than 1 year after surgery). The median survival time in this group was 1292 days with overall survival rates of 86.7%, 65.4% and 31.3% for 2, 3 and 5 years after surgery. These results confirm that recurrence has a negative impact in the prognosis.

Statistical methods for analyzing data in an illness-death multi-state model depend on the Markov assumption, which states that past and future are independent given the present state. By ignoring the disease history behavior (e.g., states previously visited and the transition times among them), these models may carry severe limitations which can make the model inappropriate. It is a fact that the future health of individuals with an early recurrence may be different from those who have been healthy for a long time. In addition, the risk of death is known to increase shortly after the recurrence, which reveals that the length of stay in the recurrence state is relevant for prognosis, thus invalidating the memoryless property of Markov processes. Accordingly, the Markov assumption can be checked by including covariates depending on the history. This 'global' test for Markovianity based on the Cox model (using time to recurrence as a covariate) reported a coefficient of negative sign for the recurrence time, according to an increased risk of death shortly after relapse (P-value = 0.154).

Since several estimators introduced in this paper are consistent regardless the Markov condition they can be used to introduce a 'local' test for the Markov condition by measuring the discrepancy in the estimates obtained from these estimators to those obtained using the Aalen-Johansen estimators (only consistent if the process is Markov). Graphical comparisons of the transition probabilities between the two approaches are reported in Figure 6. This figure depicts the discrepancy between the landmark non-Markovian estimator (LM) and the Aalen-Johansen estimator (Markovian), for $p_{12}(s,t)$ and $p_{22}(s,t)$, for $s = 365$, $s = 730$ and $s = 1095$ ($D_{ij} = p_{ij}^{\text{LM}}(s,t) - p_{ij}^{\text{AJ}}(s,t)$). The 95% pointwise confidence bands are based on simple bootstrap are also shown, revealing clear differences between the two methods in large intervals for $s = 365$. In this case, since there exists a deviation of the plot with respect to the straight line $y = 0$, one gets some evidence on the lack of Markovianity of the underlying process beyond one year after surgery. On the other hand, the plots depicted on the second and third row no not reveal evidence against the Markov assumption. In summary, these plots reveal some evidence, at least for $s = 365$, that the use of Markov-free estimators such as those

proposed in this paper are more suited to estimate the transition probabilities $p_{12}(s,t)$, $p_{13}(s,t)$, $p_{22}(s,t)$ and $p_{23}(s,t)$.
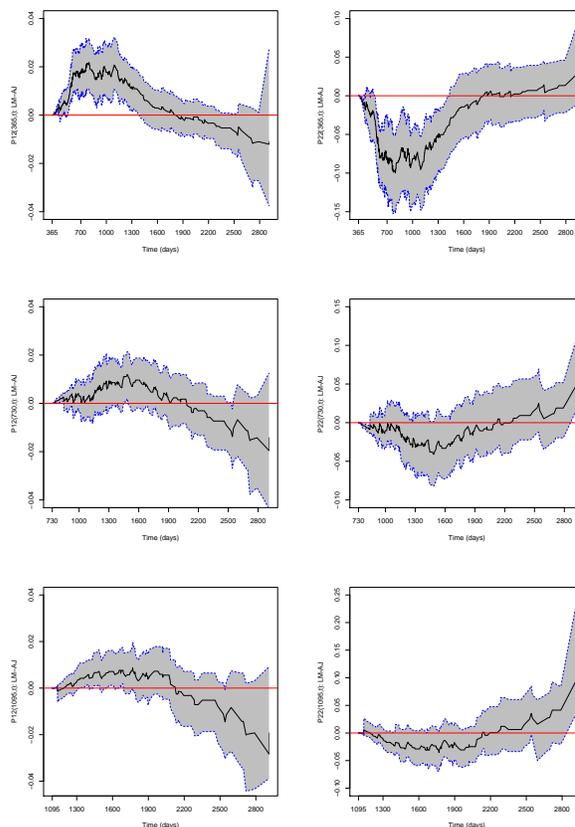

Figure 6: Local graphical test for the Markov condition, for $s = 365$ (top), $s = 730$ (middle) and $s = 1095$ (bottom). Test based on the discrepancy between the Aalen-Johansen estimator (Markovian) and the Markov-free estimator (`LM`). Colon cancer data.

With this application we are also interested in illustrating differences between the estimated transition probabilities from the estimators introduced in Section 2. These quantities can be used to obtain prediction probabilities of future events (e.g., recurrence and death in cancer studies). In Figure 7 we present, as an example, estimated transition probabilities for $p_{12}(s,t)$ and $p_{22}(s,t)$, with $s = 365$, $s = 730$ and $s = 1095$ (corresponding to 1, 2 and 3 years) for the colon cancer data, showing that a choice between the different methods makes a big difference. As shown in our simulations the estimators by [14] provide curves that are almost always below those obtained by the new estimators. This is more clear in the transition probability $p_{22}(s,t)$ for higher values of s. This is in agreement with our simulation results that suggested a

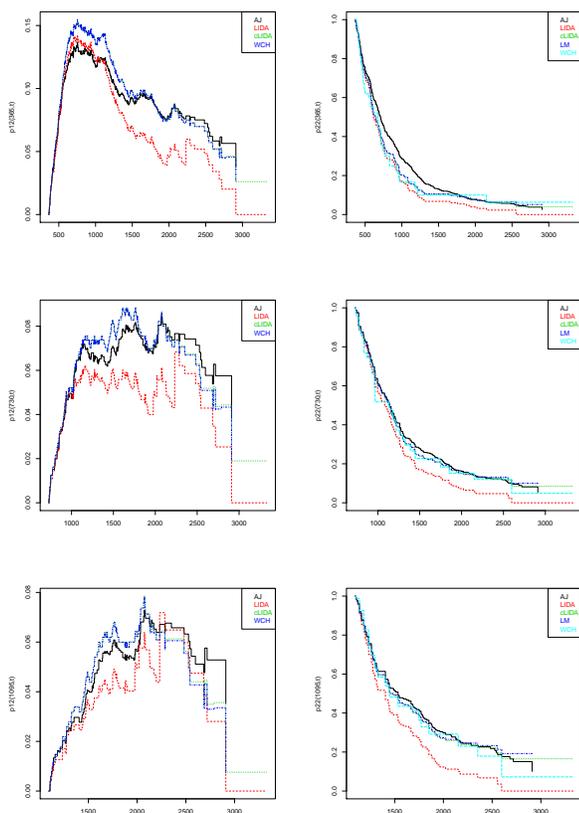systematic negative bias for the estimator by [14] (i.e. a downward biased estimator).

Figure 7: Estimated transition probabilities for $p_{12}(s,t)$ and $p_{22}(s,t)$, $s = 365$ (top), $s = 730$ (middle) and $s = 1095$ (bottom). Colon cancer data.

Since few events ('death') are observed at higher time values, consistency problems are expected at the right tail of the distribution when using the estimator by [14]. These features can be seen in all plots but especially in the figures of the transition probability $p_{22}(s,t)$. While both LM and WCH estimators decrease smoothly with time the estimator by [14] shows a sharp decrease to zero.

All plots depicted in right hand side of Figure 7 reveal a similar behavior of the LM and WCH estimators of the transition probability $p_{22}(s,t)$. These plots report the survival fraction along time, among the individuals in the recurrence state 1 year (Figure 7, top), and 2 years (Figure 7, middle) and 3 years (Figure 7, bottom) after surgery. They reveal that patients with an early recurrence have lower survival probabilities. When comparing the two Markov-free methods with the Aalen-Johansen estimator (AJ) one can observe some differences for $s = 365$ which are less evident as s increases. These discrepancies can be explained by

the failure of the Markov assumption as shown in Figure 6. Similarly, differences can also be observed between AJ and WCH estimators for the transition probability $p_{12}(s,t)$. Summarizing, it becomes clear from this application that, at least for $s = 365$, the use of Markov-free estimators such as LM and WCH are preferred over the Aalen-Johansen estimator.

5 Discussion

There has been a remarkable surge of activity lately on the topic of nonparametric estimation of transition probabilities in multi-state models. Most recent contributions on this topic are in the context of non-Markov multi-state models since the Aalen-Johansen estimator is still the preferred and standard estimator when one is confident of the Markov assumption. One recent paper has used the idea of subsampling to introduce estimators that are consistent regardless the Markov condition. In this paper we propose new estimators which are constructed using the cumulative hazard of the total time given a first time but where each observation has been weighted using the information of the first duration. Results obtained from several simulation studies conducted under different data scenarios show that the new method and the proposals introduced by [16] are quite similar providing accurate estimates.

The comparison between estimated transition probabilities is the basis to introduce a graphical local test for the Markov assumption. The new methods are based on measuring the discrepancy of the Aalen-Johansen estimator which gives consistent estimators in Markov processes, and recent approaches that do not rely on this assumption. Our simulation results indicated that the Aalen-Johansen estimator provides biased estimates if the Markov assumption does not hold. In most of these cases the use of a non-Markov estimator is preferable due to their greater accuracy. Therefore, one important issue is how to test the Markov assumption. Results reported in our data illustration reveal that the use of a local graphical test can lead to more reliable conclusions than those obtained by a global test such as the one studying Markovianity through covariates depending on history.

References:

[1] P. K. Andersen and Ø. Borgan and R. D. Gill and N. Keiding, *Statistical Models Based on Counting Processes*, Statistics in Medicine, Springer-Verlag, New York, 1993.

[2] Hougaard, P., *Multi-state Models: a Review*, Lifetime Data Analysis, 5, pp. 239–264, 1999.

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

[4] Meira-Machado, L. and Sestelo, M., *Estimation in the progressive illness-death model: A nonexhaustive review*, Biometrical Journal, 18, pp. 195–222, 2018.

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

[6] Andersen, P.K. and Esbjerj, S. and Sorensen, T.I.A., *Multistate models for bleeding episodes and mortality in liver cirrhosi*, Statistics in Medicine, 19, pp. 587–599, 2000.

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

[8] Putter, H. and Fiocco, M. and Geskus, R. B., *Non-homogeneous Markov models in the analysis of survival after breast cancer*, Tutorial in biostatistics: Competing risks and multi-state models, 26, 2007.

[9] Aalen, O. and Johansen, S., *An Empirical transition matrix for non homogeneous Markov and chains based on censored observations*, Scandinavian Journal of Statistics, 5, 141–150, 1978.

[10] Moreira, A. and de Uña-Álvarez, J. and Meira-Machado, L., *Presmoothing the Aalen-Johansen estimator in the illness-death model*, Electronical Journal of Statistics, 7, 1491–1516, 2013.

[11] Datta, S. and Satten, G.A., *Validity of the Aalen-Johansen estimators of stage occupation probabilities and Nelson Aalen integrated transition hazards for non-Markov models*, Statistics & Probability Letters, 55, 403–411, 2001.

[12] Datta, S. and Satten, G. A., *Estimation of integrated transition hazards and stage occupation probabilities for non-Markov systems under dependent censoring*, Biometrics, 58, 792–802, 2002.

[13] Glidden, D., *Robust inference for event probabilities with non-Markov event data*, Biometrics, 58, 361–368, 2002.

[14] Meira-Machado, L. and de Uña-Álvarez, J. and Cadarso-Suárez, C., *Nonparametric Estimation of Transition Probabilities in a Non-Markov Illness-Death Model*, Lifetime Data Analysis, 12, 325–344, 2006.

[15] Allignol, A. and Beyersmann, J. and Gerds, T. and Latouche, A., *Nonparametric Estimation of Transition Probabilities in a Non-Markov Illness-Death Model*, Lifetime data analysis, 20, 495–513, 2014.

[16] de Uña-Álvarez, J. and Meira-Machado, L., *Nonparametric estimation of transition probabilities in the non-Markov illness-death model: A comparative study*, Biometrics, 2, 364–375, 2015.

[17] Putter, H. and Spitoni, C., *Non-parametric estimation of transition probabilities in non-Markov multi-state models: The landmark Aalen-Johansen estimator*, Statistical Methods in Medical Research, 27, 2081–2092, 2018.

[18] Borgan Ø., *Aalen-Johansen Estimator*, Encyclopedia of Biostatistics, 1, 5–10, 1988.

[19] Kaplan, E.L. and Meier, P., *Nonparametric estimation from incomplete observations*, Journal of the American Statistical Association, 53, 457–481, 1958.

[20] Meira-Machado, L., *Smoothed landmark estimators of the transition probabilities*. SORT-Statistics and Operations Research Transactions, 40, 375–398, 2016.

[21] Soutinho, G., Meira-Machado, L. Oliveira, P., *comparison of presmoothing methods in the estimation of transition probabilities*, Communications in Statistics - Simulation and Computation, 2020. DOI: 10.1080/03610918.2020.1762895

[22] Cox, D.R., *Regression models and life tables*, Journal of the Royal Statistical Society Series B, 34, 187–220, 1972.

[23] Moertel , C. G. and Fleming , Thomas R. and Macdonald , J. S. and Haller , D. G. and Laurie, J. A. and Goodman , P. J. and Ungerleider , J. S. and Emer son , W. A. and Tormey , D. C. and Glick , J. H. and Veeder , M. H. and Mailliard , J. A., *Levamisole and Fluorouracil for Adjuvant Therapy of Resected Colon Carcinoma*, New England Journal of Medicine, 322, 352-358, 1990